

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.
OM protein - protein search, using sw model
Run on: November 7, 2005, 09:31:57 ; Search time 231 Seconds
(without alignments)
3589.672 Million cell updates/sec
Title: US-09-603-665-5
Perfect score: 10807
Sequence: 1 MTSLAQQLQRLALPOSDASL.....CQKTIQQLTVLGEPLQSYF 2144
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5
Searched: 2105692 seqs, 386760381 residues
Total number of hits satisfying chosen parameters: 2105692
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries
Database : A_Geneseq_16Dec04.*
1: Geneseqp1980s.*
2: Geneseqp1990s.*
3: Geneseqp2000s.*
4: Geneseqp2001s.*
5: Geneseqp2002s.*
6: Geneseqp2003as.*
7: Geneseqp2003bs.*
8: Geneseqp2004s.*
Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	10799	99.9	2144	4 AAB85029	Protein e
2	10799	99.9	2144	8 ADQ89800	Antagonis
3	10782	99.8	2144	8 ADS34902	Human aut
4	10222	94.6	2044	8 ADS34907	Human aut
5	9853	91.2	2036	8 ADS34905	Human aut
6	9853	91.2	2036	8 ADS34903	Human aut
7	9853	91.2	2036	8 ADS34904	Human aut
8	7756.5	71.8	1569	4 ABG15232	Novel hum
9	6124.5	56.7	1229	8 ADG66400	Novel hum
10	5026.5	46.5	1149	7 ADR08012	Novel pro
11	4506	41.7	897	8 ADG96220	T cell ac
12	4498	41.6	897	8 ADQ96218	T cell ac
13	4494	41.6	897	8 ADQ96152	T cell ac
14	3694	34.2	734	7 ADM04615	Human pro
15	2585	23.9	515	2 AAW54099	Homo sapi
16	2061	19.1	408	8 ADS34906	Human aut
17	2051.5	19.0	2096	4 ABB65242	Drosophil
18	2051.5	19.0	2096	8 ADQ89606	Antagonis
19	1779	16.5	349	8 ADP24166	PRO polyp
20	1778	16.5	349	8 ADE92729	Human pro
21	1603	14.8	325	7 ADE09054	Novel pro
22	1086	10.0	1798	6 ABJ26330	Aspergill
23	1085	10.0	1814	6 ABJ25730	Aspergill
24	1056.5	9.8	1818	5 ABP73779	Candida a
25	956	8.8	1769	6 ABR52974	Protein s

26	956	8.8	1769	7 ADK62658	Adk62658 Disease t
27	663	6.1	135	5 ABP41280	Abp41280 Human ova
28	379	3.5	77	4 AAM17368	Aam17368 Peptide #
29	379	3.5	77	4 ABB36384	Abb36384 Peptide #
30	379	3.5	77	4 AAM29882	Aam29882 Peptide #
31	379	3.5	77	4 ABB31185	Abb31185 Peptide #
32	379	3.5	77	4 ABB21738	Abb21738 Protein #
33	379	3.5	77	4 AAM69542	Aam69542 Human bon
34	379	3.5	77	4 AAM57146	Aam57146 Human bra
35	379	3.5	77	4 ABG51215	Abg51215 Human liv
36	379	3.5	77	4 AAM05057	Aam05057 Peptide #
37	379	3.5	77	5 ABG39167	Abg39167 Human pep
38	309.5	2.9	2228	7 ABR61599	Abri61599 Human gol
39	309.5	2.9	2230	6 ABU07445	Abu07445 Protein d
40	309.5	2.9	2230	7 ABR61600	Abri61600 Human gol
41	308	2.9	2250	7 ABR61601	Abri61601 Human gol
42	308	2.9	2252	7 ABR61602	Abri61602 Human gol
43	295	2.7	4952	8 ADK15818	Adk15818 Human ABC
44	295	2.7	4958	8 ADK15814	Adk15814 Human ABC
45	295	2.7	5058	8 ADK15793	Adk15793 Human ABC

ALIGNMENTS

RESULT 1
AAB85029
ID AAB85029 standard; protein; 2144 AA.
XX
AC AAB85029;
XX
DT 06-AUG-2001 (first entry)
XX
DE Protein encoded by BAP28 cDNA consisting of exons 1 to 45.
XX
KW BAP28; prostate; tumour; cancer; diagnostic; genetic analysis.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 1694
FT Misc-difference /label= Ser or Asn
FT Misc-difference 1854
FT Misc-difference /label= Ala or Val
FT Misc-difference 1967
FT Misc-difference /label= Asp or Asn
FT Misc-difference 2017
FT Misc-difference /label= Gly or Glu
XX
WO200100669-A2.
XX
O4-JAN-2001.
XX
23-JUN-2000; 2000WO-IB001183.
XX
25-JUN-1999; 99US-0141323P.
XX
18-JAN-2000; 2000US-0176880P.
XX
(GEST) GENSET.
XX
PI Barry C, Bougueleret L, Chumakov I, Cohen-Akenine A;
XX
WPI; 2001-367032/38.
XX
N-PSDB; AAF83909, AAF83910.
XX
PT New BAP28 polynucleotides and polypeptides overexpressed in prostate
PT cancer cells for diagnostic prostate tumors, e.g. by hybridization or
PT polymerase chain reaction assays.
XX
PS Claim 14; Page 297-304; 349pp; English.
XX
CC The invention is directed to BAP28 polypeptides, BAP28 polynucleotide
CC sequences and regulatory region located at the 3' and 5' ends of the

CC BAP28 coding region. The BAP28 polypeptides can be expressed by standard
CC recombinant methodology. BAP28 polynucleotides and polypeptides have been
CC found to be over expressed in prostate tumour cells, therefore levels of
CC BAP28 expression and/or activity may be assayed (e.g. by polymerase chain
CC reaction (PCR)) to diagnose patient suffering from or susceptible to
CC prostate cancer. Antibodies specific for the BAP28 polypeptides are
CC useful as diagnostic reagents. Biallelic markers of the BAP28 gene are
CC useful in genetic analysis. The present sequence represents a protein
CC encoded by a first cDNA sequence of the BAP28 gene consisting of the
CC exons 1 to 45
XX
SQ Sequence 2144 AA;

Query Match		99.9%;	Score 10799;	DB 4;	Length 2144;	
Best Local Similarity		100.0%;	Pred. No. 0;			
Matches 2144;		Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;	
Qy	1	MTSLAQQLQRLALPOS	DASLLSRDEVASLLFPK	EAATIDRDTAF	AGCTGLELLGIDP 60	
Db	1	MTSLAQQLQRLALPOS	DASLLSRDEVASLLFPK	EAATIDRDTAF	AGCTGLELLGIDP 60	
Qy	61	SFEQPEAPLFSQ	LAKTLERSVOTKAVNKOLDENISLFI	HLSPYFLLPKPAQKCLEWLIHR	120	
Db	61	SFEQPEAPLFSQ	LAKTLERSVOTKAVNKOLDENISLFI	HLSPYFLLPKPAQKCLEWLIHR	120	
Qy	121	PHIHLNQDSL	LIACVLPYHETRI FVRVIQ	LLKINNSKRWFWLLPVKQSGVPLAKGT	LI 180	
Db	121	PHIHLNQDSL	LIACVLPYHETRI FVRVIQ	LLKINNSKRWFWLLPVKQSGVPLAKGT	LI 180	
Qy	181	HCYKDLGFMDF	TCSLVTKSVKFAEYPGSSAOLRV	LAFYASTIVSALVA	AEVDSDNI 240	
Db	181	HCYKDLGFMDF	TCSLVTKSVKFAEYPGSSAOLRV	LAFYASTIVSALVA	AEVDSDNI 240	
Qy	241	KLFPYIQKLS	PSLPYRAATYMIICQISVKVT	MENTFVNSLASQIIKTLT	KIPSLIKDG 300	
Db	241	KLFPYIQKLS	PSLPYRAATYMIICQISVKVT	MENTFVNSLASQIIKTLT	KIPSLIKDG 300	
Qy	301	LSCLIVLLQ	RQRPESIGKPPFPHLCNVPDL	ITILHGISETYDVSP	PLRLMPLHVLVSI 360	
Db	301	LSCLIVLLQ	RQRPESIGKPPFPHLCNVPDL	ITILHGISETYDVSP	PLRLMPLHVLVSI 360	
Qy	361	HVTGETEGMD	GQIYKHEALITKISLAKN	LDHLASLLFPEYISYSSOE	MDSNKS 420	
Db	361	HVTGETEGMD	GQIYKHEALITKISLAKN	LDHLASLLFPEYISYSSOE	MDSNKS 420	
Qy	421	LNQFPLRL	LESKYPTRLDVVLEHKEIADL	KKQELFHQFVSLSTSGKYQ	FLADSD 480	
Db	421	LNQFPLRL	LESKYPTRLDVVLEHKEIADL	KKQELFHQFVSLSTSGKYQ	FLADSD 480	
Qy	481	TSMLSLNHP	LAPVRILAMNHLKIMKTSKEGVDESFI	KEAVLARIGD	DNIDVLSA 540	
Db	481	TSMLSLNHP	LAPVRILAMNHLKIMKTSKEGVDESFI	KEAVLARIGD	DNIDVLSA 540	
Qy	541	FEIFKEHFS	SEVTISNLLNLFQRAELSKNGEWYVLKIAADILI	KEEILSENDOLSNQV	600	
Db	541	FEIFKEHFS	SEVTISNLLNLFQRAELSKNGEWYVLKIAADILI	KEEILSENDOLSNQV	600	
Qy	601	VCLLPFWIN	NDTSAEMKIAIYLSKSGICSLHPLLRGWEALEN	VIKSTPGKLI	660	
Db	601	VCLLPFWIN	NDTSAEMKIAIYLSKSGICSLHPLLRGWEALEN	VIKSTPGKLI	660	
Qy	661	NQKMIELL	ADNINLGDPSMLKMWEDLISVGEESFNLKQVTF	HFVILSVLVS	CCSSLKE 720	
Db	661	NQKMIELL	ADNINLGDPSMLKMWEDLISVGEESFNLKQVTF	HFVILSVLVS	CCSSLKE 720	
Qy	721	THEPFAIR	PSLLQKKIKLESVITAVE	IPSEWHIELMDRGI	PVELWAHYVEELNSTOR 780	
Db	721	THEPFAIR	PSLLQKKIKLESVITAVE	IPSEWHIELMDRGI	PVELWAHYVEELNSTOR 780	
Qy	781	VAVEDSV	FLVFLSKKFIYAL	KAPSPFKGDIWNP	PEQLKEDSDRYLHLLIGLFEM	MLNGA 840
Db	781	VAVEDSV	FLVFLSKKFIYAL	KAPSPFKGDIWNP	PEQLKEDSDRYLHLLIGLFEM	MLNGA 840

QY	841	DAVHFRVLMKLF	IKVHLEDV	FQLFKFSVLW	TGSSLSNPLNCS	VKTVL	QALYVGCAM	900			
DB	841	DAVHFRVLMKLF	IKVHLEDV	FQLFKFSVLW	TGSSLSNPLNCS	VKTVL	QALYVGCAM	900			
QY	901	LSSQKTQCKHOLAS	ISSPVV	TSLINLGS	PKVEVRRAAI	QCLQAL	SGVASPFYLIIDHLI	960			
DB	901	LSSQKTQCKHOLAS	ISSPVV	TSLINLGS	PKVEVRRAAI	QCLQAL	SGVASPFYLIIDHLI	960			
QY	961	SKABEITS	DAAYVQD	LATLFEEL	OREKLLKSHQK	LSLKNLAS	CVYSCPSYIAKDLMK	1020			
DB	961	SKABEITS	DAAYVQD	LATLFEEL	OREKLLKSHQK	LSLKNLAS	CVYSCPSYIAKDLMK	1020			
QY	1021	VLQGVNG	EMVLSQ	LLPMABOLLEKI	QKEPTAVL	KDEAMVHL	TLTGKNEFVSLLNEDPK	1080			
DB	1021	VLQGVNG	EMVLSQ	LLPMABOLLEKI	QKEPTAVL	KDEAMVHL	TLTGKNEFVSLLNEDPK	1080			
QY	1081	SLDIFIKAVHTTKEL	YAGMPTIQT	IALEKITK	FPFAAISDEKVQ	OKLLRMLFD	LLVNCN	1140			
DB	1081	SLDIFIKAVHTTKEL	YAGMPTIQT	IALEKITK	FPFAAISDEKVQ	OKLLRMLFD	LLVNCN	1140			
QY	1141	SHCAQTVSSVFKGIS	SVNAEQVRI	HELEPPDKAK	PLGTVOOKRQ	MOOKSODLES	QVQEVG	1200			
DB	1141	SHCAQTVSSVFKGIS	SVNAEQVRI	HELEPPDKAK	PLGTVOOKRQ	MOOKSODLES	QVQEVG	1200			
QY	1201	GSYQWV	TLILELQ	HKKLRSPQ	ILVPTLFNLLSR	CLPLPQEOGNMEY	TKQLILSCLL	1260			
DB	1201	GSYQWV	TLILELQ	HKKLRSPQ	ILVPTLFNLLSR	CLPLPQEOGNMEY	TKQLILSCLL	1260			
QY	1261	NI	CQKLSPDGK	GI	PKDILDEEKENVELIVQ	CIRLSEMPQTHHALLLGL	TVAGIFPDKVL	1320			
DB	1261	NI	CQKLSPDGK	GI	PKDILDEEKENVELIVQ	CIRLSEMPQTHHALLLGL	TVAGIFPDKVL	1320			
QY	1321	HNIMSI	FTFMGANVR	LLDDTVS	QVINKTVKMPALIOS	DSGDSIEVSRNVEEIVVKII	1380				
DB	1321	HNIMSI	FTFMGANVR	LLDDTVS	QVINKTVKMPALIOS	DSGDSIEVSRNVEEIVVKII	1380				
QY	1381	SVFVDALPHVPHRR	RLPILVOL	DTLGA	BKFLWILLIL	FEQYVTKTVL	AAAYGKDAL	1440			
DB	1381	SVFVDALPHVPHRR	RLPILVOL	DTLGA	BKFLWILLIL	FEQYVTKTVL	AAAYGKDAL	1440			
QY	1441	EADTEFWFVSC	CEFSVQHQIQSL	MNIQYLLKLP	PEKEETIPKAVSFNKSQ	EEMLQVF	1500				
DB	1441	EADTEFWFVSC	CEFSVQHQIQSL	MNIQYLLKLP	PEKEETIPKAVSFNKSQ	EEMLQVF	1500				
QY	1501	NVETHTSKQLRH	FKFLSVFMSQ	LLSSNNFLKKV	VS	GGPEILKGLERLL	TVLGYISA	1560			
DB	1501	NVETHTSKQLRH	FKFLSVFMSQ	LLSSNNFLKKV	VS	GGPEILKGLERLL	TVLGYISA	1560			
QY	1561	VAQSMERNADKL	TVKFWRALLSKAYD	LLDKVNALLPTETFI	PVIRGLVGNPLPS	VRRKAL	1620				
DB	1561	VAQSMERNADKL	TVKFWRALLSKAYD	LLDKVNALLPTETFI	PVIRGLVGNPLPS	VRRKAL	1620				
QY	1621	DLANNKLQONI	SWKKTIV	TRFLKVPDLLAI	VORKKKEGEEBQA	INRQALVTLKLLCN	1680				
DB	1621	DLANNKLQONI	SWKKTIV	TRFLKVPDLLAI	VORKKKEGEEBQA	INRQALVTLKLLCN	1680				
QY	1681	FGAENPDDPPV	PLTAVKLIA	APERKEEKNVLSALLCIAE	VTSLEALIPOLPSLMPSL	1740					
DB	1681	FGAENPDDPPV	PLTAVKLIA	APERKEEKNVLSALLCIAE	VTSLEALIPOLPSLMPSL	1740					
QY	1741	LTTMKNTSEL	VSSVYLLSALAALQKVETL	PHFISPYLEGILISQVILHEKITSEMGAS	1800						
DB	1741	LTTMKNTSEL	VSSVYLLSALAALQKVETL	PHFISPYLEGILISQVILHEKITSEMGAS	1800						
QY	1801	QANIRL	TS	LKKTAT	TAPR	VLLPAIKTKTYKO	IEKWKVNHMGPFMSILOEHIGXMKKEEL	1860			
DB	1801	QANIRL	TS	LKKTAT	TAPR	VLLPAIKTKTYKO	IEKWKVNHMGPFMSILOEHIGXMKKEEL	1860			
QY	1861	TSHQS	QLTAFF	LEALD	FRAG	HSEND	LEEVGKTENCIIDCLVAMVVKLSVTTPPLPKLF	1920			
DB	1861	TSHQS	QLTAFF	LEALD	FRAG	HSEND	LEEVGKTENCIIDCLVAMVVKLSVTTPPLPKLF	1920			
QY	1921	DWAKTE	DAPK	DRILL	TFYN	LAD	CI	AEKLGFLTFAGHLVKP	FADTLTXQVNI	SKTDEAFFD	1980

Db 1921 DWAKTEDADPKRLTLFTYNLADCAIEAKGLFTLFAHGLVKFPADTLXQVNIKTDEAFD 1980
Qy 1981 SENDPEKCCLLQFILNCLYKIFLFTQHFISKERAXALMPLVDOLNRLGEEKFQER 2040
Db 1981 SENDPEKCCLLQFILNCLYKIFLFTQHFISKERAXALMPLVDOLNRLGEEKFQER 2040
Qy 2041 VTKHLIPICIAQSVAMADSLWKPLNYQILLTKTROSSPKVRFAALITVLALAEKLENYI 2100
Db 2041 VTKHLIPICIAQSVAMADSLWKPLNYQILLTKTROSSPKVRFAALITVLALAEKLENYI 2100
Qy 2101 VLLPESIPPLAEIMEDECEVEHQCKTIQOLETVLGEPLQSVF 2144
Db 2101 VLLPESIPPLAEIMEDECEVEHQCKTIQOLETVLGEPLQSVF 2144

RESULT 2

ADQ89800

ID ADQ89800 standard; protein; 2144 AA.

XX AC

XX AC

XX AC

DT 21-OCT-2004 (first entry)

XX DE

XX DE Antagonist of cell cycle progression polypeptide #115.

XX KW

XX KW Cytostatic; cancer; cell division cycle; mitosis; meiosis;

XX KW cell cycle progression.

XX OS

XX OS Homo sapiens.

XX PN

XX PN WO2004063362-A2.

XX PD

XX PD 29-JUL-2004.

XX PF

XX PF 31-DEC-2003; 2003WO-CB005635.

XX PR

XX PR 10-JAN-2003; 2003US-0439123P.

XX PR

XX PR 06-MAY-2003; 2003US-0468402P.

XX PA

XX PA (CYCL-) CYCLACEL LTD.

XX PI

XX PI Glover D, Bell G, Frenz L, Midgley C;

XX PX

XX PX WPI; 2004-544089/52.

XX DR

XX DR N-PSDB; ADQ89799.

XX PT

XX PT New cell cycle progression genes and proteins for modulating cell cycle

XX PT progression in cells, for preventing, treating or diagnosing cell

XX PT proliferative diseases (e.g. cancer) or for identifying modulators of

XX PT mitosis or meiosis.

XX PS

XX PS Claim 2; SEQ ID NO 230; 461pp; English.

XX CC

XX CC The present invention relates to a polynucleotide for preventing,

XX CC treating or diagnosing a disease in an individual. The composition or the

XX CC polypeptide, polynucleotide or RNA precursor, or antibody is useful for

XX CC diagnosing, preventing or treating diseases (e.g. cell proliferative

XX CC diseases such as cancer) in an individual. These may also be used for

XX CC identifying substances capable of binding to or modulating the function

XX CC of the polypeptide, capable of affecting the function of the

XX CC corresponding gene, or capable of inhibiting the cell division cycle or

XX CC cell cycle progression, preferably mitosis and/or meiosis. The present

XX CC sequence represents an antagonist of cell cycle progression protein

XX CC sequence.

XX SX

XX SX Sequence 2144 AA;

XX SX

XX SX Query Match

XX SX Best Local Similarity 99.9%; Score 10799; DB 8; Length 2144;

XX SX Matches 2140; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

XX SX

XX SX 1 MTSLAQQLQRLALPQSDASLLSRDEVASLLFDPKEAATIDRTAPAGCTGLELLGIDP 60

Db 1 MTSLAQQLQRLALPQSDASLLSRDEVASLLFDPKEAATIDRTAPAGCTGLELLGIDP 60
Qy 61 SPEQFEAPLFSOLAKTLERSVCTKAVNKOLDENISLFIHLSPYFLPKPAQCLEWLIHR 120
Db 61 SPEQFEAPLFSOLAKTLERSVCTKAVNKOLDENISLFIHLSPYFLPKPAQCLEWLIHR 120
Qy 121 FHIHLYNQDSLACVLPYHETRIFVRVIQLLKINNSKHFWFWLLPVKQSGVPLAKGTLLIT 180
Db 121 FHIHLYNQDSLACVLPYHETRIFVRVIQLLKINNSKHFWFWLLPVKQSGVPLAKGTLLIT 180
Qy 181 HCYKOLGFMDFICSLVTKSVKFAEPYPGSSAQLRLVLLAFYASTIVSALVAEDVSNIIA 240
Db 181 HCYKOLGFMDFICSLVTKSVKFAEPYPGSSAQLRLVLLAFYASTIVSALVAEDVSNIIA 240
Qy 241 KLFPYIQKGLKSSLPDYRAATYMIICQISVKVTMENTFVNSLASOIIKTTLKIPSLIKOG 300
Db 241 KLFPYIQKGLKSSLPDYRAATYMIICQISVKVTMENTFVNSLASOIIKTTLKIPSLIKOG 300
Qy 301 LSLCLIVLLQKQKPESLGKPPHLCNVPDLITILHGISTYDVSPLLRYMLPHLVVSIH 360
Db 301 LSLCLIVLLQKQKPESLGKPPHLCNVPDLITILHGISTYDVSPLLRYMLPHLVVSIH 360
Qy 361 HVTGEETEGMDQIYKRHLLEAILTKISLKNLDHLLASLLFEEYISYSSQEBMSNKVSL 420
Db 361 HVTGEETEGMDQIYKRHLLEAILTKISLKNLDHLLASLLFEEYISYSSQEBMSNKVSL 420
Qy 421 LNEQFLPLIRLLESKYPRITDVLBEHLKEIADLKKQELFHQFVLSITSGGKYQFVLADSD 480
Db 421 LNEQFLPLIRLLESKYPRITDVLBEHLKEIADLKKQELFHQFVLSITSGGKYQFVLADSD 480
Qy 481 TSLMLSLNHP LAPVRILAMNHLKIMTKSKEGVDESFIKEAVLARLGDDNIDVLSAISA 540
Db 481 TSLMLSLNHP LAPVRILAMNHLKIMTKSKEGVDESFIKEAVLARLGDDNIDVLSAISA 540
Qy 541 FEIKFHEFSSEVTISNLLNLFORAELSKNGEYEVLTAAADILIKBEILSENDQLSNQV 600
Db 541 FEIKFHEFSSEVTISNLLNLFORAELSKNGEYEVLTAAADILIKBEILSENDQLSNQV 600
Qy 601 VCLLPFVVINDDTESAEWKIAIYLSKGI CSLHPLLRGWEALENVIKSTPKGLIGVA 660
Db 601 VCLLPFVVINDDTESAEWKIAIYLSKGI CSLHPLLRGWEALENVIKSTPKGLIGVA 660
Qy 661 NQMIELLDNINLGDPSMLKMWEDLISVGEESFNKQKVTFFHVLVSVSCSSLKE 720
Db 661 NQMIELLDNINLGDPSMLKMWEDLISVGEESFNKQKVTFFHVLVSVSCSSLKE 720
Qy 721 THFPPAIRVFSLLQKIKKLESVITAVTIPSEWHIELMDRGI PVELWAHYVEELNSTOR 780
Db 721 THFPPAIRVFSLLQKIKKLESVITAVTIPSEWHIELMDRGI PVELWAHYVEELNSTOR 780
Qy 781 VAVEDSVFLVSLKFIYALKAPKSPFKGDIWNPEQLKEDSRDYLHLLIGLFEMMLNCA 840
Db 781 VAVEDSVFLVSLKFIYALKAPKSPFKGDIWNPEQLKEDSRDYLHLLIGLFEMMLNCA 840
Qy 841 DAVHFRVLMKLFIKVHLEDVFLQFKFCSVLMTYSSLSNPLNCSVKTVLQALYVGCAM 900
Db 841 DAVHFRVLMKLFIKVHLEDVFLQFKFCSVLMTYSSLSNPLNCSVKTVLQALYVGCAM 900
Qy 901 LSSQKQCKHOLASISSPVVTSLLINLGSVPKVRRAAIQCLQALSGVSPYLLIDHLI 960
Db 901 LSSQKQCKHOLASISSPVVTSLLINLGSVPKVRRAAIQCLQALSGVSPYLLIDHLI 960
Qy 961 SKABEITSDAAVVIQDLATLFEELQREKKLKHOKLSETLKNLLSCVYSCPSYIAKDLMK 1020
Db 961 SKABEITSDAAVVIQDLATLFEELQREKKLKHOKLSETLKNLLSCVYSCPSYIAKDLMK 1020
Qy 1021 VLQVNGEMVLSQLPMAEQLEKIQEPTAVLKDEAMVHLTLGKYNEFSVSLNEDPK 1080
Db 1021 VLQVNGEMVLSQLPMAEQLEKIQEPTAVLKDEAMVHLTLGKYNEFSVSLNEDPK 1080
Qy 1081 SIDIFIKAHVHTTKELYAGMPTIOITALEKTIKTPFAAISDEKVOQKLLMFLDVLVNCN 1140

Db 1081 SLIDIFIKAVHTTKELYAGMPTIQTALKEKITKPFPAIISDEKVQOKLRLMFLDLLVNCKN 1140
Qy 1141 SHCAQTVSVFSGISVNAEQVRIELEPPDKAKPLGTVOQKRQKMOQKKSODLESQVBVG 1200
Db 1141 SHCAQTVSVFSGISVNAEQVRIELEPPDKAKPLGTVOQKRQKMOQKKSODLESQVBVG 1200
Qy 1201 GSYMQRVTLIILELQHKKKLRSQILVPTLFNLLSRCLPLPQEQGNMEYTKQLILSCLL 1260
Db 1201 GSYMQRVTLIILELQHKKKLRSQILVPTLFNLLSRCLPLPQEQGNMEYTKQLILSCLL 1260
Qy 1261 NICQKLSPDGKIPKDIIDBEKFNVELIVQCIRLSEMQTHHALLILGTVAGIPDKVL 1320
Db 1261 NICQKLSPDGKIPKDIIDBEKFNVELIVQCIRLSEMQTHHALLILGTVAGIPDKVL 1320
Qy 1321 HNIMISFTFMGANVRLDDTYSFQVINKTKVWIPALIQSDSGDSIEVSRNVBEIIVWKII 1380
Db 1321 HNIMISFTFMGANVRLDDTYSFQVINKTKVWIPALIQSDSGDSIEVSRNVBEIIVWKII 1380
Qy 1381 SVFVDALPHVPEHRLPLVLQVLDTLGABKFLWILLILFEQYVTKTVLAAAYGEKDAIL 1440
Db 1381 SVFVDALPHVPEHRLPLVLQVLDTLGABKFLWILLILFEQYVTKTVLAAAYGEKDAIL 1440
Qy 1441 EADTFWFVSCCFVSQHQIQSLMILQVLLKLPKEKETIPKAVSFNKSESQEMLOVF 1500
Db 1441 EADTFWFVSCCFVSQHQIQSLMILQVLLKLPKEKETIPKAVSFNKSESQEMLOVF 1500
Qy 1501 NVETHTSKQLRHFKFLVSFMSQLLSSNNFLKXVSGGPEILKGLERLLETVLGYISA 1560
Db 1501 NVETHTSKQLRHFKFLVSFMSQLLSSNNFLKXVSGGPEILKGLERLLETVLGYISA 1560
Qy 1561 VAQSMERNADKLTVKFWRALSKAYDLDKVNALLPTETTFIPVIRGLVGNPLPSVRRKAL 1620
Db 1561 VAQSMERNADKLTVKFWRALSKAYDLDKVNALLPTETTFIPVIRGLVGNPLPSVRRKAL 1620
Qy 1621 DLLNNKLOQNIISWKTTIVTRFLKLPDILLAIYORKKKEGEERQAINRQTALYTLKLCCKN 1680
Db 1621 DLLNNKLOQNIISWKTTIVTRFLKLPDILLAIYORKKKEGEERQAINRQTALYTLKLCCKN 1680
Qy 1681 FGAENPDPPVPLXLTAVKLIAPERKEKNVLSALICIAEVTSTLEALAIPOPLSIMPSL 1740
Db 1681 FGAENPDPPVPLXLTAVKLIAPERKEKNVLSALICIAEVTSTLEALAIPOPLSIMPSL 1740
Qy 1741 LTTMKNTSELVSEVYLLSALAALQKVETLPHFISPYLEGILSQVILHEKITSEMGSSAS 1800
Db 1741 LTTMKNTSELVSEVYLLSALAALQKVETLPHFISPYLEGILSQVILHEKITSEMGSSAS 1800
Qy 1801 QANIRLTSUKTLATTLAPRVLLPAIKTKYKQIEKNKNHMGPFMSILOEHIGXMKKEEL 1860
Db 1801 QANIRLTSUKTLATTLAPRVLLPAIKTKYKQIEKNKNHMGPFMSILOEHIGXMKKEEL 1860
Qy 1861 TSHQSOLTAFLEALDFAQHSENDEEVGKTENCIIICLVAMVVKLSEVTFRPLFFKLF 1920
Db 1861 TSHQSOLTAFLEALDFAQHSENDEEVGKTENCIIICLVAMVVKLSEVTFRPLFFKLF 1920
Qy 1921 DWAKTEDAPKORLLTFYNIADCTAEKLGFTLTFAGHLVKPFDATLXQVNI SKTDRAFD 1980
Db 1921 DWAKTEDAPKORLLTFYNIADCTAEKLGFTLTFAGHLVKPFDATLXQVNI SKTDRAFD 1980
Qy 1981 SENDPEKCCILLQFILNCLYKIFLFTDQHPISKERAXALMPLVDQLENRLGGEERQOER 2040
Db 1981 SENDPEKCCILLQFILNCLYKIFLFTDQHPISKERAXALMPLVDQLENRLGGEERQOER 2040
Qy 2041 VTKHLIPICIAQFSVAMADDSLWKPLNVIQILLKTRDSSPKVRFAALITVLALAEKLENYI 2100
Db 2041 VTKHLIPICIAQFSVAMADDSLWKPLNVIQILLKTRDSSPKVRFAALITVLALAEKLENYI 2100
Qy 2101 VLLPESIPFLAELMEDCEBEVHCQCKTIQOLETVLGEPLQSYF 2144
Db 2101 VLLPESIPFLAELMEDCEBEVHCQCKTIQOLETVLGEPLQSYF 2144

ID AD34902 standard; protein; 2144 AA.
XX ADS34902;
AC 16-DEC-2004 (first entry)
XX Human autoimmune disease-related protein - SEQ ID 116.
DE single nucleotide polymorphism detection; SNP detection;
KW rheumatoid arthritis; type 1 diabetes; multiple sclerosis;
KW systemic lupus erythematosus; inflammatory bowel disease; psoriasis;
KW thyroiditis; celiac disease; pernicious anaemia; asthma; vitiligo;
KW glomerulonephritis; Grave's disease; myocarditis; Sjogren's disease;
KW primary systemic vasculitis.
XX Homo sapiens.
OS WO2004083403-A2.
XX 30-SEP-2004.
XX 18-MAR-2004; 2004WO-US008461.
XX 18-MAR-2003; 2003US-0455444P.
XX 25-APR-2003; 2003US-0465241P.
XX (APPL-) APPLERA CORP.
XX Cargill M, Begovich AB, Alexander HC;
XX WPI; 2004-728480/71.
XX N-PSDB; ADS34819.
XX New isolated nucleic acid molecule comprises at least 8 contiguous
PT nucleotides where one of the nucleotides is a single nucleotide
PT polymorphism (SNP), useful for diagnosing or treating autoimmune
PT diseases, e.g. rheumatoid arthritis.
XX Claim 12; SEQ ID NO 116; 123pp; English.
XX The invention comprises amino acid and coding sequences containing
CC genetic polymorphisms associated with an altered risk of developing an
CC autoimmune disease (e.g. rheumatoid arthritis). The invention further
CC comprises a method of identifying an individual that has an altered risk
CC of developing an autoimmune disease, comprising detecting a single
CC nucleotide polymorphism (SNP) in a nucleic acid of the invention. The DNA
CC and protein sequences of the invention are useful for diagnosing and
CC treating autoimmune diseases, such as: rheumatoid arthritis, type 1
CC diabetes, multiple sclerosis, systemic lupus erythematosus, inflammatory
CC bowel diseases, psoriasis, thyroiditis, celiac disease, pernicious
CC anaemia, asthma, vitiligo, glomerulonephritis, Grave's disease, The
CC myocarditis, Sjogren's disease, or primary systemic vasculitis. The
CC present amino acid sequence represents a human autoimmune disease-related
CC protein of the invention. NOTE: The present sequence is not shown in the
CC specification, but has been retrieved from the WIPO website.
XX Sequence 2144 AA;
Qy Query Match 99.8%; Score 10782; DB 8; Length 2144;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 2137; Conservative 1; Mismatches 6; Indels 0; Gaps 0;
Qy 1 MTSLAQQLQRLALPQSDASLLSRDEVASILFQPKAATDRDTAFAGTCGEBELGIDP 60
Db 1 MTSLAQQLQRLALPQSDASLLSRDEVASILFQPKAATDRDTAFAGTCGEBELGIDP 60
Qy 61 SPEQFEAPLFSQAKTLERSVQTKAVNKOLDENISLFIHLSPYFLPKPAKCLEWLIHR 120
Db 61 SPEQFEAPLFSQAKTLERSVQTKAVNKOLDENISLFIHLSPYFLPKPAKCLEWLIHR 120
Qy 121 FTHILYNQDSLTIACVLPYHETRI FVRVIOQLKINNKHFWFLLPVKQSGVPLAKGTLIT 180
Db 121 FTHILYNQDSLTIACVLPYHETRI FVRVIOQLKINNKHFWFLLPVKQSGVPLAKGTLIT 180

181 HCYKOLGFMDFICSLVTKSVKFAEYPGSSAQLRVLAFYASTIVSALVAEDVSDNIIA 240
181 HCYKDLGFMDFICSLVTKSVKFAEYPGSSAQLRVLAFYASTIVSALVAEDVSDNIIA 240
241 KLPFYLOKGLKSSLPDYRAATYMIICQISVKVMTMENTFVNSLASQIIKTITKPSIKDG 300
241 KLPFYLOKGLKSSLPDYRAATYMIICQISVKVMTMENTFVNSLASQIIKTITKPSIKDG 300
301 LSLCLIVLLQKQKPESLGKPFPHLCNVDPDLITILHGISYDYVSPILRYMPLHLVSIH 360
301 LSLCLIVLLQKQKPESLGKPFPHLCNVDPDLITILHGISYDYVSPILRYMPLHLVSIH 360
361 HVTGEETEGMDGQIYKXHLBAIITKISLKNLHLLASLLFEBYISYSSQEEEMDSNKVSL 420
361 HVTGEETEGMDGQIYKXHLBAIITKISLKNLHLLASLLFEBYISYSSQEEEMDSNKVSL 420
421 LNEQFLPLRLLESKYPRTLDDVLEHLKEIADLKQELPHQFVSLSTSGKYQFLADSD 480
421 LNEQFLPLRLLESKYPRTLDDVLEHLKEIADLKQELPHQFVSLSTSGKYQFLADSD 480
481 TSLMLSLNHPLAPVRILAMNHLKKIMKTSKEGVDESFKEAVLARLGDDNIDVLSAISA 540
481 TSLMLSLNHPLAPVRILAMNHLKKIMKTSKEGVDESFKEAVLARLGDDNIDVLSAISA 540
541 FEIFKHFSESVTISNLLNLFORAELSKNGEYEVILKIAADILIKBEIILSENQOLSNQVV 600
541 FEIFKHFSESVTISNLLNLFORAELSKNGEYEVILKIAADILIKBEIILSENQOLSNQVV 600
601 VCLLPFWVNNDDTESAEMKAIYLSKSGICSLHPLLRGWEELALENVIKSTPKGLIGVA 660
601 VCLLPFWVNNDDTESAEMKAIYLSKSGICSLHPLLRGWEELALENVIKSTPKGLIGVA 660
661 NQKMIELLADNINLGPSSMLKMWEDLISVGEESFNLLKQKVTFFHVLVSVSCSSLKE 720
661 NQKMIELLADNINLGPSSMLKMWEDLISVGEESFNLLKQKVTFFHVLVSVSCSSLKE 720
721 THPPFAIRVFLSLQKKIKKLESVITAVEIPSEWHIELMLDRGIPVELWAHYVEELNSTOR 780
721 THPPFAIRVFLSLQKKIKKLESVITAVEIPSEWHIELMLDRGIPVELWAHYVEELNSTOR 780
781 VAVEDSVFLVSLKKEIYALKAPKSPKGDIMNPEQLKEDSDYILHLIGLFEMMLNGA 840
781 VAVEDSVFLVSLKKEIYALKAPKSPKGDIMNPEQLKEDSDYILHLIGLFEMMLNGA 840
841 DAHFRVLMKLFIKVHLEDVFOQLFKFCVSLWTYGSLSNPLNCSVKTVLQTOALYVGCAM 900
841 DAHFRVLMKLFIKVHLEDVFOQLFKFCVSLWTYGSLSNPLNCSVKTVLQTOALYVGCAM 900
901 LSSQKTCQKHQLASISSPVVTSLLINLGPSPKVEVRRAAIOCLQALSGVASPPYLIIDHLI 960
901 LSSQKTCQKHQLASISSPVVTSLLINLGPSPKVEVRRAAIOCLQALSGVASPPYLIIDHLI 960
961 SKAEBITSDAAVYIQLATLFEELQREKKLKHQKSETLKNLLSCVYSCPSYIAKDLMK 1020
961 SKAEBITSDAAVYIQLATLFEELQREKKLKHQKSETLKNLLSCVYSCPSYIAKDLMK 1020
1021 VLQGVNGEMVLSOLLPMWQLLEKIQKEPTAVLKDEAMVHLTLGKYNEFSVSLNEDPK 1080
1021 VLQGVNGEMVLSOLLPMWQLLEKIQKEPTAVLKDEAMVHLTLGKYNEFSVSLNEDPK 1080
1081 SLIDIFKAVHTTKELYAGMPTIQTITALEKITTTPFFAAISDEKVOQKLLRMLFDLLVNCN 1140
1081 SLIDIFKAVHTTKELYAGMPTIQTITALEKITTTPFFAAISDEKVOQKLLRMLFDLLVNCN 1140
1141 SHCAQTSSVFKGISVNAQVRIELEPPDKAKPLGTVOQKRRQKMOOKSODLESVQEVG 1200
1141 SHCAQTSSVFKGISVNAQVRIELEPPDKAKPLGTVOQKRRQKMOOKSODLESVQEVG 1200
1201 GSYWQVRLIILELLOHKKLRSPQILVPTLFLNLLSRCLBPLPOEQGNMEYTKQILISCLL 1260
1201 GSYWQVRLIILELLOHKKLRSPQILVPTLFLNLLSRCLBPLPOEQGNMEYTKQILISCLL 1260

1261 NICQKLSPDGKIPKIDILDEEKENVELIVQICIRLSEMPQTHHALLLGTVAGIFPDKVL 1320
1261 NICQKLSPDGKIPKIDILDEEKENVELIVQICIRLSEMPQTHHALLLGTVAGIFPDKVL 1320
1321 HNIMSIPTFMGANVMBLDDTYSQVINKTVKMVIPALIQSDSGDSIEVSNNVEIIVVKII 1380
1321 HNIMSIPTFMGANVMBLDDTYSQVINKTVKMVIPALIQSDSGDSIEVSNNVEIIVVKII 1380
1381 SVEFDALPHVPEHRRPLILVQLVDTLGAEKFWILLILIFEQVYVTKVTLAAAVGEKDAIL 1440
1381 SVEFDALPHVPEHRRPLILVQLVDTLGAEKFWILLILIFEQVYVTKVTLAAAVGEKDAIL 1440
1441 EADTEFWFVCCBFSVQHOIQSLMNTLOYLKLLPBEKEETIPKAVSFNKSESQEBMLQVF 1500
1441 EADTEFWFVCCBFSVQHOIQSLMNTLOYLKLLPBEKEETIPKAVSFNKSESQEBMLQVF 1500
1501 NVETHTSKOLRHFKFLSVFMSOLLSSNNFLKVVESGGPEILKGLLEERLLETVLGVISA 1560
1501 NVETHTSKOLRHFKFLSVFMSOLLSSNNFLKVVESGGPEILKGLLEERLLETVLGVISA 1560
1561 VAQSMERNADKLTVKFWRALLSKAYDLDDKNALLPTETFIPVIRGLVGNPLPSVRKAL 1620
1561 VAQSMERNADKLTVKFWRALLSKAYDLDDKNALLPTETFIPVIRGLVGNPLPSVRKAL 1620
1621 DLLNNKLQONI SWKKTIVTRFLKLVDPDLAIIVORKKKEGEEQAINRQTALYTLKLLCKN 1680
1621 DLLNNKLQONI SWKKTIVTRFLKLVDPDLAIIVORKKKEGEEQAINRQTALYTLKLLCKN 1680
1681 FGAENPDPPVVLXTAVKLIAPERKEEKVLSGALLCIAEVTSTLEALAIPOPLSLMPSL 1740
1681 FGAENPDPPVVLXTAVKLIAPERKEEKVLSGALLCIAEVTSTLEALAIPOPLSLMPSL 1740
1741 LTTMKNTSELVSEVYLLSALAALOKVETLPHFTISPYLEGILSOVHLEKITSEMGAS 1800
1741 LTTMKNTSELVSEVYLLSALAALOKVETLPHFTISPYLEGILSOVHLEKITSEMGAS 1800
1801 QANIRLTSUKKTATTAPRVLLPAIKKTYKQIEKNWKNHMGPFMSILOEHIGMKKEEL 1860
1801 QANIRLTSUKKTATTAPRVLLPAIKKTYKQIEKNWKNHMGPFMSILOEHIGMKKEEL 1860
1861 TSHQSOLTAFFLEALDFAHSENDELSEVGKTENCIICLCLVAMVVKLSEVTPRPLPKLF 1920
1861 TSHQSOLTAFFLEALDFAHSENDELSEVGKTENCIICLCLVAMVVKLSEVTPRPLPKLF 1920
1921 DWAKTEDAPKRLTFYNLADCIAEKLKGLFTLFAHGLVKFPADTLKQVNI SKTDEAFD 1980
1921 DWAKTEDAPKRLTFYNLADCIAEKLKGLFTLFAHGLVKFPADTLKQVNI SKTDEAFD 1980
1981 SENDPEKCCLLQFILNCLYKIFLFDTHFISKERAXALMPLVDQLENRLGGEKQFER 2040
1981 SENDPEKCCLLQFILNCLYKIFLFDTHFISKERAXALMPLVDQLENRLGGEKQFER 2040
2041 VTKHLIPCIAOFSVAMADDSLMKPNYQILLKTRDSSPKVRFAALTIVLALAEKLENYI 2100
2041 VTKHLIPCIAOFSVAMADDSLMKPNYQILLKTRDSSPKVRFAALTIVLALAEKLENYI 2100
2101 VLLPESIPPLAELMEDECEVEHQCKTIQOLETVLGEPLQSYF 2144
2101 VLLPESIPPLAELMEDECEVEHQCKTIQOLETVLGEPLQSYF 2144

RESULT 4

ADS34907

ID ADS34907 standard; protein; 2044 AA.

AC ADS34907;

XX

DT 16-DEC-2004 (first entry)

XX

DE Human autoimmune disease-related protein - SEQ ID 121.

XX

KW single nucleotide polymorphism detection; SNP detection;

KW rheumatoid arthritis; type 1 diabetes; multiple sclerosis;

KW systemic lupus erythematosus; inflammatory bowel disease; psoriasis;
KW thyroiditis; celiac disease; pernicious anaemia; asthma; vitiligo;
KW glomerulonephritis; Grave's disease; myocarditis; Sjogren's disease;
XX primary systemic vasculitis.
OS Homo sapiens.
XX WO2004083403-A2.
XX 30-SEP-2004.
XX 18-MAR-2004; 2004WO-US008461.
XX 18-MAR-2003; 2003US-0455444P.
XX 25-APR-2003; 2003US-0465241P.
XX (APPL-) APPLERA CORP.
XX Cargill M, Begovich AB, Alexander HC;
PI WPI; 2004-728480/71.
XX N-PSDB; AD834824.
XX New isolated nucleic acid molecule comprises at least 8 contiguous
PT nucleotides where one of the nucleotides is a single nucleotide
PT polymorphism (SNP), useful for diagnosing or treating autoimmune
PT diseases, e.g. rheumatoid arthritis.
XX Claim 12; SEQ ID NO 121; 123pp; English.
PS The invention comprises amino acid and coding sequences containing
XX genetic polymorphisms associated with an altered risk of developing an
CC autoimmune disease (e.g. rheumatoid arthritis). The invention further
CC comprises a method of identifying an individual that has an altered risk
CC of developing an autoimmune disease, comprising detecting a single
CC nucleotide polymorphism (SNP) in a nucleic acid of the invention. The DNA
CC and protein sequences of the invention are useful for diagnosing and
CC treating autoimmune diseases, such as: rheumatoid arthritis, type 1
CC diabetes, multiple sclerosis, systemic lupus erythematosus, inflammatory
CC bowel diseases, psoriasis, thyroiditis, celiac disease, pernicious
CC anaemia, asthma, vitiligo, glomerulonephritis, Grave's disease,
CC myocarditis, Sjogren's disease, or primary systemic vasculitis. The
CC present amino acid sequence represents a human autoimmune disease-related
CC protein of the invention. NOTE: The present sequence is not shown in the
CC specification, but has been retrieved from the WIPO website.
XX Sequence 2044 AA;
SQ
Query Match 94.6%; Score 10222; DB 8; Length 2044;
Best Local Similarity 95.0%; Pred. No. 0;
Matches 2037; Conservative 1; Mismatches 6; Indels 100; Gaps 2;
QY 1 MTSIAQQLQRLALPQSDASLLSRDEVASLLFDPEKATIDRDTAFPAIGCTGTELLGIDP 60
DB 1 MTSIAQQLQRLALPQSDASLLSRDEVASLLFDPEKATIDRDTAFPAIGCTGTELLGIDP 60
QY 61 SFEQFEAPLFSQAKTLERSVQTKAVNKQLDENISLFLHLSPYFLKPAQKCLEWLIHR 120
DB 61 SFEQFEAPLFSQAKTLERSVQTKAVNKQLDENISLFLHLSPYFLKPAQKCLEWLIHR 120
QY 121 FHIHLNQDSLIAACVLPYHETRI FVRVI QLLKTNNSKHRFWLLPVKQSGVPLAKGTLIT 180
DB 121 FHIHLNQDSLIAACVLPYHETRI FVRVI QLLKTNNSKHRFWLLPVKQSGVPLAKGTLIT 180
QY 181 HCYKDLGFMDFICSLVTKSVKVAEYPGSSAQLRVLIAFYASTIVSALVAEVDSDNIIA 240
DB 181 HCYKDLGFMDFICSLVTKSVKVAEYPGSSAQLRVLIAFYASTIVSALVAEVDSDNIIA 240
QY 241 KLPFYIQGLKSSLPDYRAATYMIICQISVKVTMENTFVNSLASQIIKTLTKIPSLIKDG 300
DB 241 KLPFYIQGLKSSLPDYRAATYMIICQISVKVTMENTFVNSLASQIIKTLTKIPSLIKDG 300
QY 301 LSLGLIVLQRPESLGGKPPHLCNVDPDLITILHGISETYDVSPLLYMPLPHLVVSIH 360

DB LSLGLIVLQRPESLGGKPPHLCNVDPDLITILHGISETYDVSPLLYMPLPHLVVSIH 360
QY 361 HVTGETEGMDQIYKRHLEAILTKISLKNLHDLLASLLPEEYISYSQSEMDSNKVS 420
DB 361 HVTGETEGMDQIYKRHLEAILTKISLKNLHDLLASLLPEEYISYSQSEMDSNKVS 420
QY 421 LNEQFLPIRLLESKYPRTLDVVLEEHKEADLKKOELFHQFVLSLSSGGKYQFLAUSD 480
DB 421 LNEQFLPIRLLESKYPRTLDVVLEEHKEADLKKOELFHQFVLSLSSGGKYQFLAUSD 480
QY 481 TSLMLSLNHLAPVRILAMNHLKIMTKTSKEGVDESFTKEAVLARLGDNDIDVLSAISA 540
DB 475 -----EGVDESFTKEAVLARLGDNDIDVLSAISA 504
QY 541 FEIFKEHFSSEVTISNLLNLFORAELSKNGEYEVVLKIAADILIKEILSENDQLSNQV 600
DB 505 FEIFKEHFSSEVTISNLLNLFORAELSKNGEYEVVLKIAADILIKEILSENDQLSNQV 564
QY 601 VCLLPFVWINDDTESAEKIAIYLSKSGICSLHPLLRGWEAEALENVIKSTPKGLIGVA 660
DB 565 VCLLPFVWINDDTESAEKIAIYLSKSGICSLHPLLRGWEAEALENVIKSTPKGLIGVA 624
QY 661 NOKMIELLADNINILGDPSSMLKMVEDLISVGBEESFNLKQKVTFFHVLVSVCCSLKE 720
DB 625 NOKMIELLADNINILGDPSSMLKM----- 647
QY 721 THFFFAIRVFSLLQKKIKKLESVITAVEIPSEWHIELMLDRGIPVELMAHYVEELNSTQR 780
DB 648 -----BIPSEWHIELMLDRGIPVELMAHYVEELNSTQR 680
QY 781 VAVEDSVFLVPSLKFIYALKAPKFPKGDINWNEPOLKEDSDVHLHLIGIFEMWLNCA 840
DB 681 VAVEDSVFLVPSLKFIYALKAPKFPKGDINWNEPOLKEDSDVHLHLIGIFEMWLNCA 740
QY 841 DAVHFRVLMKLFKVHLEDFVOLFVCSVLWTVGSSLSNPLNCVKTVLQTAQYVYGCAM 900
DB 741 DAVHFRVLMKLFKVHLEDFVOLFVCSVLWTVGSSLSNPLNCVKTVLQTAQYVYGCAM 800
QY 901 LSSQKTQCKHQIASISSPVVTSLLNLGSPVKEVRRAAITQCLQALSGVASPFYLIIDHLI 960
DB 801 LSSQKTQCKHQIASISSPVVTSLLNLGSPVKEVRRAAITQCLQALSGVASPFYLIIDHLI 860
QY 961 SKAEBITSDAAVIODLATLFEELQREKKLKHOKLSETLKNLSCVYSPSVIAKDLMK 1020
DB 861 SKAEBITSDAAVIODLATLFEELQREKKLKHOKLSETLKNLSCVYSPSVIAKDLMK 920
QY 1021 VLQGVNGENVLSQLLPMAEQLLLEKIQKPTAVLKDEAMVHLTLGKYNEFSVSLNEDPK 1080
DB 921 VLQGVNGENVLSQLLPMAEQLLLEKIQKPTAVLKDEAMVHLTLGKYNEFSVSLNEDPK 980
QY 1081 SLDIFIKAVHTTKELVAGMPTTIQITALEKIKTPFFAAISDEKVVQQLRMLFDLLVNCN 1140
DB 981 SLDIFIKAVHTTKELVAGMPTTIQITALEKIKTPFFAAISDEKVVQQLRMLFDLLVNCN 1040
QY 1141 SHCAQTVSSVFGKISVNAEQVRIELEPPDKAPLGTVOQKRBQKMOOKSODLESVQEVG 1200
DB 1041 SHCAQTVSSVFGKISVNAEQVRIELEPPDKAPLGTVOQKRBQKMOOKSODLESVQEVG 1100
QY 1201 GSVQWRVTLILELLOHKKKLRSPQILVPTLFLNLLSRCLBPLPQEQGNMEYTKQLILSCLL 1260
DB 1101 GSVQWRVTLILELLOHKKKLRSPQILVPTLFLNLLSRCLBPLPQEQGNMEYTKQLILSCLL 1160
QY 1261 NICQKLSPOGKIPKDIIDEEKFNVELIVOCIRLSEMPOTHHALLLGLTAVAGIPDKVL 1320
DB 1161 NICQKLSPOGKIPKDIIDEEKFNVELIVOCIRLSEMPOTHHALLLGLTAVAGIPDKVL 1220
QY 1321 HNIMSIFTFMGANVMDLDDTYSFOVINKTVKVIIPALIOSDSDGDSIEVSRNVEEIVVKII 1380
DB 1221 HNIMSIFTFMGANVMDLDDTYSFOVINKTVKVIIPALIOSDSDGDSIEVSRNVEEIVVKII 1280
QY 1381 SVFVDALPHVPEHRRPLIILVQLVDTLGAEBKFWLILLLPEQYVTKVTLAAAYGEKDAIL 1440

Db 1281 SVFDALPHVPHRRPILVQVDTLGAKEFLMILLILLPEQVVTKTVLAACGCKDAIL 1340
Qy 1441 EADTEFWFSCCFVSQHQIQSLMNLQYLLKLPKEKEETIPKAVSFNKSSESQEMLQVF 1500
Db 1341 EADTEFWFSCCFVSQHQIQSLMNLQYLLKLPKEKEETIPKAVSFNKSSESQEMLQVF 1400
Qy 1501 NVETHTSKOLRHPKFLSVSPMSQQLSSNNFLKVVESGGEPIILKGLERLLETVLGVISA 1560
Db 1401 NVETHTSKOLRHPKFLSVSPMSQQLSSNNFLKVVESGGEPIILKGLERLLETVLGVISA 1460
Qy 1561 VAQSMERNADKLVKFWALLSKAYDLDKVNALLPTETPIVIRGLVGNPLPSVRKAL 1620
Db 1461 VAQSMERNADKLVKFWALLSKAYDLDKVNALLPTETPIVIRGLVGNPLPSVRKAL 1520
Qy 1621 DLNNKLQNNISWKKTIVTRFLKVPDLLAIVQRKKKEGEEQAINRQTALYTLKLLCKN 1680
Db 1521 DLNNKLQNNISWKKTIVTRFLKVPDLLAIVQRKKKEGEEQAINRQTALYTLKLLCKN 1580
Qy 1681 FGAENDDPFPVPLXTAVKLIAPERKEKNVLSGALLCIAEVTSTLEALAIPQLPSLPSL 1740
Db 1581 FGAENDDPFPVPLXTAVKLIAPERKEKNVLSGALLCIAEVTSTLEALAIPQLPSLPSL 1640
Qy 1741 LTTMKNTSELVSSEVLLSALAALQKVETLPHFISPYLEGILSOVHLEKITSEMGAS 1800
Db 1641 LTTMKNTSELVSSEVLLSALAALQKVETLPHFISPYLEGILSOVHLEKITSEMGAS 1700
Qy 1801 QANIRLTSKKTATLAPRVLLPAIKKTYKQIEKNKMHGPFMSILOEHIGXMKKEEL 1860
Db 1701 QANIRLTSKKTATLAPRVLLPAIKKTYKQIEKNKMHGPFMSILOEHIGXMKKEEL 1760
Qy 1861 TSHQSOLTAFFLEALDFAHQSENDELEVKTENCIDCLVAMVKLSEVTRPLPFKLF 1920
Db 1761 TSHQSOLTAFFLEALDFAHQSENDELEVKTENCIDCLVAMVKLSEVTRPLPFKLF 1820
Qy 1921 DWAKTEDAPKDRLLTFYNLADCTAEKLGKFTLFGHVKPFPADTLXOVNISKTDFAFD 1980
Db 1821 DWAKTEDAPKDRLLTFYNLADCTAEKLGKFTLFGHVKPFPADTLXOVNISKTDFAFD 1880
Qy 1981 SENDPEKCCLLQFILNCLYKIFLFTQHFISKERAXALMPLVDQLENLGEKEKFQR 2040
Db 1881 SENDPEKCCLLQFILNCLYKIFLFTQHFISKERAXALMPLVDQLENLGEKEKFQR 1940
Qy 2041 VTKHLIPCTAQSVMADDSLWKPLNYQILLKTRDSSPKVRFAALITVLAELAKENYI 2100
Db 1941 VTKHLIPCTAQSVMADDSLWKPLNYQILLKTRDSSPKVRFAALITVLAELAKENYI 2000
Qy 2101 VLLPESIPFLAELMEDECEVEHOCQKTIQOLETVLGEPLQSYF 2144
Db 2001 VLLPESIPFLAELMEDECEVEHOCQKTIQOLETVLGEPLQSYF 2044

RESULT 5

ADS34905

ID ADS34905 standard; protein; 2036 AA.

XX ADS34905;

AC ADS34905;

XX 16-DEC-2004 (first entry)

DT Human autoimmune disease-related protein - SEQ ID 119.

DE single nucleotide polymorphism detection; SNP detection;

XX rheumatoid arthritis; type 1 diabetes; multiple sclerosis;

KW systemic lupus erythematosus; inflammatory bowel disease; psoriasis;

KW thyroiditis; celiac disease; pernicious anaemia; asthma; vitiligo;

KW glomerulonephritis; Grave's disease; myocarditis; Sjogren's disease;

KW Primary systemic vasculitis.

XX Homo sapiens.

OS WO2004083403-A2.

XX 30-SEP-2004.

PN

XX

PD

XX

PF

PR

PR

PR

XX

PA

XX

PI

XX

DR

DR

XX

PT

PT

PT

XX

PS

XX

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

18-MAR-2004; 2004WO-US008461.

18-MAR-2003; 2003US-0455444P.

25-APR-2003; 2003US-0465241P.

(APPL-) APPLERA CORP.

Cargill M, Begovich AB, Alexander HC;

WPI; 2004-728480/71.

N-PSDB; ADS34822.

New isolated nucleic acid molecule comprises at least 8 contiguous

nucleotides where one of the nucleotides is a single nucleotide

polymorphism (SNP), useful for diagnosing or treating autoimmune

diseases, e.g. rheumatoid arthritis.

Claim 12; SEQ ID NO 119; 123pp; English.

The invention comprises amino acid and coding sequences containing

genetic polymorphisms associated with an altered risk of developing an

autoimmune disease (e.g. rheumatoid arthritis). The invention further

comprises a method of identifying an individual that has an altered risk

of developing an autoimmune disease, comprising detecting a single

nucleotide polymorphism (SNP) in a nucleic acid of the invention. The DNA

and protein sequences of the invention are useful for diagnosing and

treating autoimmune diseases, such as: rheumatoid arthritis, type 1

diabetes, multiple sclerosis, systemic lupus erythematosus, inflammatory

bowel diseases, psoriasis, thyroiditis, celiac disease, pernicious

anaemia, asthma, vitiligo, glomerulonephritis, Grave's disease,

myocarditis, Sjogren's disease, or primary systemic vasculitis. The

present amino acid sequence represents a human autoimmune disease-related

protein of the invention. NOTE: The present sequence is not shown in the

specification, but has been retrieved from the WIPO website.

Sequence 2036 AA;

Query Match 91.2%; Score 9853; DB 8; Length 2036;

Best Local Similarity 91.5%; Pred. No. 0;

Matches 1978; Conservative 14; Mismatches 26; Indels 144; Gaps 7;

Qy 1 MTSLAQQLQRLALPOSASLLSRDEVASLLFPFKEAATIDRTAFAGTGTLEELGIDP 60

Db 1 MTSLAQQLQRLALPOSASLLSRDEVASLLFPFKEAATIDRTAFAGTGTLEELGIDP 60

Qy 61 SPEQPEAPLFSOLAKTLERSVCTKAVNKOLDENISLFLHLSPYFLKPAQKCLEWLIHR 120

Db 61 SPEQPEAPLFSOLAKTLERSVCTKAVNKOLDENISLFLHLSPYFLKPAQKCLEWLIHR 120

Qy 121 FHIHLNODSLIACVLPYHETRI FVRVIQLLKNNKSHRWFLLPVKQSGVPLAKGTLIT 180

Db 121 FHIHLNODSLIACVLPYHETRI FVRVIQLLKNNKSHRWFLLPVKQSGVPLAKGTLIT 180

Qy 181 HCYKDLGFMDFICSLVTKSVKVFABYPGSSAQLRVLLAFYASTIVSALVAEDVSNIIA 240

Db 181 HCYKDLGFMDFICSLVTKSVKVFABYPGSSAQLRVLLAFYASTIVSALVAEDVSNIIA 240

Qy 241 KLFPYIQLKSSLPDYRAATYMIICQISVKVTM---ENTFVNSLASQ-1IKTLTKIPS- 295

Db 241 KLFPYIQLKSSLPDYRAATYMIICQISVKVTM---YLTGRTETQVSLRHQNDQDSLPDQGVKLLDSAPAE 289

Qy 296 -----LIKDGLSCLIVLLQKPKESLKGKPPHLCNVPDLITILHGISETYD 342

Db 296 TEAREPWEKVCVTIELRNGAS-----QRPFPHLCNVPDLITILHGISETYD 334

Qy 343 VSPLLRYMLPHLVSIHHVTGTEGMDGQIYKQHLEAILTKISLKNLHLLASLLE 402

Db 343 VSPLLRYMLPHLVSIHHVTGTEGMDGQIYKQHLEAILTKISLKNLHLLASLLE 402

Qy 403 EYISYSSQEMDSNKNVSLNEQFLPLRLLESKYPTLDVLEELKEIADLKQELFQ 462

Db 395 EYISYSSQEMDSNKNVSLNEQFLPLRLLESKYPTLDVLEELKEIADLKQELFQ 454

PA (APPL-) APPLERA CORP.
XX Cargill M, Begovich AB, Alexander HC;
XX WPI; 2004-728480/71.
DR N-PSDB; ADS34820.
XX
XX New isolated nucleic acid molecule comprises at least 8 contiguous
PT nucleotides where one of the nucleotides is a single nucleotide
PT polymorphism (SNP), useful for diagnosing or treating autoimmune
PT diseases, e.g. rheumatoid arthritis.
XX
XX Claim 12; SEQ ID NO 117; 123pp; English.
PS
PS The invention comprises amino acid and coding sequences containing
CC genetic polymorphisms associated with an altered risk of developing an
CC autoimmune disease (e.g. rheumatoid arthritis). The invention further
CC comprises a method of identifying an individual that has an altered risk
CC of developing an autoimmune disease, comprising detecting a single
CC nucleotide polymorphism (SNP) in a nucleic acid of the invention. The DNA
CC and protein sequences of the invention are useful for diagnosing and
CC treating autoimmune diseases, such as: rheumatoid arthritis, type 1
CC diabetes, multiple sclerosis, systemic lupus erythematosus, inflammatory
CC bowel diseases, psoriasis, thyroiditis, celiac disease, pernicious
CC anaemia, asthma, vitiligo, glomerulonephritis, Grave's disease,
CC myocarditis, Sjogren's disease, or primary systemic vasculitis. The
CC present amino acid sequence represents a human autoimmune disease-related
CC protein of the invention. NOTE: The present sequence is not shown in the
CC specification, but has been retrieved from the WIPO website.
XX
XX Sequence 2036 AA;
SQ

Query Match 91.2%; Score 9853; DB 8; Length 2036;
Best Local Similarity 91.5%; Pred. No. 0;
Matches 1978; Conservative 14; Mismatches 26; Indels 144; Gaps 7;

QY 1 MTSLAQQLRALPQSDASLLSRDEVASLLDFDPEKAATIDRTAFAGCTGLEELGIDP 60
DB 1 MTSLAQQLRALPQSDASLLSRDEVASLLDFDPEKAATIDRTAFAGCTGLEELGIDP 60

QY 61 SPEQFEAPLFSQAKTLERSVQTKAVNKOLDENISLFLHLSPYFLLPKPAQCLEWLIHR 120
DB 61 SPEQFEAPLFSQAKTLERSVQTKAVNKOLDENISLFLHLSPYFLLPKPAQCLEWLIHR 120

QY 121 FHIHLNQDSLACVLPYHETRI FVRVIO LKINNSKRWFWMLLPVKQSGVPLAKGTLIT 180
DB 121 FHIHLNQDSLACVLPYHETRI FVRVIO LKINNSKRWFWMLLPVKQSGVPLAKGTLIT 180

QY 181 HCYKDLGFWDFICSLVTKSVKPAEYPGSSAQLRVLLAFYASTIVSALVAEDVSDNIIA 240
DB 181 HCYKDLGFWDFICSLVTKSVKPAEYPGSSAQLRVLLAFYASTIVSALVAEDVSDNIIA 240

QY 241 KLFPYIKQLGKLSPLDYAATYMIICQISVKVTM---ENTFVNSLASQ-IIKTLTKIPS- 295
DB 241 KLFPYIKQLS- - - - -YLTGRTEQTSLSRHQNIQDLSDFDQGVKLLDSAPAE 289

QY 296 - - - - -LIKDGLSCLIVLLQROKPSGLKPPPHLCNVPDLITILHGISETYD 342
DB 296 TEAREPEWKCVTIELNGAS- - - - -QRPFPPLCNVPDLITILHGISETYD 334

QY 343 VSPLLRYMLPHLVVSTIIHHVTGTEGMDGOIKYKRLHLEAILTKISLKNLHLLASLLE 402
DB 343 VSPLLRYMLPHLVVSTIIHHVTGTEGMDGOIKYKRLHLEAILTKISLKNLHLLASLLE 402

QY 403 EYISYSSQEMDSNKSVLNNEOPLIRLLESKYPTLDVLEELHLEKADLKQBLFHQ 462
DB 403 EYISYSSQEMDSNKSVLNNEOPLIRLLESKYPTLDVLEELHLEKADLKQBLFHQ 462

QY 463 FVSLSTSGKYQFLADSDTSLMLSLNHLAPVRIILAMNHLKIMTKTSKEGVDESFIKEAV 522
DB 463 FVSLSTSGKYQ- - - - -EGVDESFIKEAV 478

QY 523 LARLGDDNIDVLSAISAFEIPEKHSSEVTISNLLNLFQRAELSKNGEYEVYLKIAADI 582

DB 479 LARLGDDNIDVLSAISAFEIPEKHSSEVTISNLLNLFQRAELSKNGEYEVYLKIAADI 538
QY 583 LIKEEILSENDOISNOVVCLLPVFINNDTDESAMKIAIYLSKSGICSLHPLLRGWE 642
DB 539 LIKEEILSENDOISNOVVCLLPVFINNDTDESAMKIAIYLSKSGICSLHPLLRGWE 598
QY 643 ALENVIKSTKPGKLIQVANQKMIELADININLGDPSMLKMMVBDLISVGBEESFNLKQV 702
DB 599 ALENVIKSTKPGKLIQVANQKMIELADININLGDPSMLKM- - - - - 639
QY 703 THFVILSVLSCSSSLKETHFPFAIRVFSLLQKKIKKLESVITAVIPSEWHIELMDRG 762
DB 640 - - - - -BIPSEWHIELMDRG 654

QY 763 IPVLMWAHYVEELNSTORVAVEDSVFLSKKFIYALKAPKSPFKGDIWNPEOLKEDS 822
DB 655 IPVLMWAHYVEELNSTORVAVEDSVFLSKKFIYALKAPKSPFKGDIWNPEOLKEDS 714

QY 823 RDLVHLLILGFEMMLNGADAVHFRVLMKLFIVKHLBDFVQLFKFCSVLMTYSSLSNPLN 882
DB 715 RDLVHLLILGFEMMLNGADAVHFRVLMKLFIVKHLBDFVQLFKFCSVLMTYSSLSNPLN 774

QY 883 CSVKTVLQTOALYVGCMLSSQKQCKHOLASISSPVTSLINLGSPKVEVRRRAIOCL 942
DB 775 CSVKTVLQTOALYVGCMLSSQKQCKHOLASISSPVTSLINLGSPKVEVRRRAIOCL 834

QY 943 QALSGVASPFYLIIDHLISKABEITSDAAYVQDLATLFEELQREKKLSHOKSLSTLKN 1002
DB 835 QALSGVASPFYLIIDHLISKABEITSDAAYVQDLATLFEELQREKKLSHOKSLSTLKN 894

QY 1003 LLSCVYSCPSYIAKOLMKVLQGVNGEMVLSQLLPMABQLEKIQKEPTAVLKDEAMVHL 1062
DB 895 LLSCVYSCPSYIAKOLMKVLQGVNGEMVLSQLLPMABQLEKIQKEPTAVLKDEAMVHL 954

QY 1063 TLGKYNESVSLNEDPKSLDIFIKAVHTTKELYAGMPTIOITALEKIKTPFAAISDEK 1122
DB 955 TLGKYNESVSLNEDPKSLDIFIKAVHTTKELYAGMPTIOITALEKIKTPFAAISDEK 1014

QY 1123 VQKLLRMLFDLLVNCNCKSHCAQTSSVPKGI SVNAEQVRIELEPPDKAPKLTGVQOKR 1182
DB 1015 VQKLLRMLFDLLVNCNCKSHCAQTSSVPKGI SVNAEQVRIELEPPDKAPKLTGVQOKR 1074

QY 1183 QKQOKKKSODLESVEGSGSYWQVVTLLLELLOHKKKLSPQILVPTLFNLSRCLPLP 1242
DB 1075 QKQOKKKSODLESVEGSGSYWQVVTLLLELLOHKKKLSPQILVPTLFNLSRCLPLP 1134

QY 1243 QKQOKKKSODLESVEGSGSYWQVVTLLLELLOHKKKLSPQILVPTLFNLSRCLPLP 1302
DB 1135 QKQOKKKSODLESVEGSGSYWQVVTLLLELLOHKKKLSPQILVPTLFNLSRCLPLP 1194

QY 1303 HALLLLGTVAGIFPDKVLHNSIIFPMGANVRLDDTYSFOVINTKVMVTPALITQSDS 1362
DB 1195 HALLLLGTVAGIFPDKVLHNSIIFPMGANVRLDDTYSFOVINTKVMVTPALITQSDS 1254

QY 1363 GDSIEVSRNVEIIVKIIISVFDALPHVPEHRRRLPILVOLVDTLGAEKFLWILLIFEQ 1422
DB 1255 GDSIEVSRNVEIIVKIIISVFDALPHVPEHRRRLPILVOLVDTLGAEKFLWILLIFEQ 1314

QY 1423 VYTKTVLAAAYGEKDAILEADTEFPWSCCEFSVOHQISLMNILOYLKLPKEEBETIP 1482
DB 1315 VYTKTVLAAAYGEKDAILEADTEFPWSCCEFSVOHQISLMNILOYLKLPKEEBETIP 1374

QY 1483 KAVSFNKSSESQEMQLQVFNVEHTSKQLRHFPKLSVFSFMSQLSSNNFLKVVESGGEPI 1542
DB 1375 KAVSFNKSSESQEMQLQVFNVEHTSKQLRHFPKLSVFSFMSQLSSNNFLKVVESGGEPI 1434

QY 1543 LKGLBERLLETVLGIVISAQSMERNADKLTVKFWRALLSKAYDILLDKVNALLPTTFTIP 1602
DB 1435 LKGLBERLLETVLGIVISAQSMERNADKLTVKFWRALLSKAYDILLDKVNALLPTTFTIP 1494

QY 1603 VTRGLVGNPLPSVRKALDILLNNKLOONI SWKKTITVTRFLKLVPLDLATVORKKGESE 1662

Db 1495 VIRGLVGNLPVRRKALDNLNKKQQNISWKKTVITRFLKLVDPDLLAIVQRKKGEGBE 1554
Qy 1663 QAINRQTALYTLKLLCKNFGAENPDFFVPLVATVKLIAPERKEKNVLSALLCIAEVT 1722
Db 1555 QAINRQTALYTLKLLCKNFGAENPDFFVPLVATVKLIAPERKEKNVLSALLCIAEVT 1614
Qy 1723 STLEALAIQPLSLMPSLLTTWKNTSELVSSEVYLLSALAALQKVETLPHFISPLYEG 1782
Db 1615 STLEALAIQPLSLMPSLLTTWKNTSELVSSEVYLLSALAALQKVETLPHFISPLYEG 1674
Qy 1783 LSQVHLEKITSEMGSASQANIRLTSLKKTTLATTLAPRVLPAIKKTKYQIEKNWNHMG 1842
Db 1675 LSQVHLEKITSEMGSASQANIRLTSLKKTTLATTLAPRVLPAIKKTKYQIEKNWNHMG 1734
Qy 1843 PFMISILQEHIGMKKEELTSHQSQTAFLEALDFAHQSENDEBEVGTENCIIIDCLVA 1902
Db 1735 PFMISILQEHIGMKKEELTSHQSQTAFLEALDFAHQSENDEBEVGTENCIIIDCLVA 1794
Qy 1903 MVVKLSEVTRFPLFFKLFDWAKTEDAPKORLLTFFYNLADCIABKLGTLFAGHLVKPF 1962
Db 1795 MVVKLSEVTRFPLFFKLFDWAKTEDAPKORLLTFFYNLADCIABKLGTLFAGHLVKPF 1854
Qy 1963 ADTLXOVNISKTDAPFDSNDPEKCLLLQFLNCLYKIFLFDTOHFISKERAXALMMP 2022
Db 1855 ADTLXOVNISKTDAPFDSNDPEKCLLLQFLNCLYKIFLFDTOHFISKERAXALMMP 1914
Qy 2023 LVDQLENRLGGEKFOERTVTKHLIPCIAQFSVAMADDSLWKPLNYQILLKTRDSGPKVRF 2082
Db 1915 LVDQLENRLGGEKFOERTVTKHLIPCIAQFSVAMADDSLWKPLNYQILLKTRDSGPKVRF 1974
Qy 2083 AALITVLAALAEKLENYIVLLPESIPFLAELMEDECEVEHQCKTIQOLETVLGEPLOS 2142
Db 1975 AALITVLAALAEKLENYIVLLPESIPFLAELMEDECEVEHQCKTIQOLETVLGEPLOS 2034
Qy 2143 YF 2144
Db 2035 YF 2036

RESULT 7
ADS34904
ID ADS34904 standard; protein; 2036 AA.
XX ADS34904;
AC ADS34904;
XX
DT 16-DEC-2004 (first entry)
XX Human autoimmune disease-related protein - SEQ ID 118.
XX single nucleotide polymorphism detection; SNP detection;
KW rheumatoid arthritis; type 1 diabetes; multiple sclerosis;
KW systemic lupus erythematosus; inflammatory bowel disease; psoriasis;
KW thyroiditis; celiac disease; pernicious anaemia; asthma; vitiligo;
KW glomerulonephritis; Grave's disease; myocarditis; Sjogren's disease;
KW primary systemic vasculitis.

XX Homo sapiens.
XX WO2004083403-A2.
XX 30-SEP-2004.
XX 18-MAR-2004; 2004WO-US008461.
XX 18-MAR-2003; 2003US-0455444P.
PR 25-APR-2003; 2003US-0465241P.
XX (APPL-) APPLERA CORP.
XX Cargill M, Begovich AB, Alexander HC;
PI WPI; 2004-728480/71.
XX N-PSDB; ADS34821.

XX New isolated nucleic acid molecule comprises at least 8 contiguous
PT nucleotides where one of the nucleotides is a single nucleotide
PT polymorphism (SNP), useful for diagnosing or treating autoimmune
PT diseases, e.g. rheumatoid arthritis.
XX
PS Claim 12; SEQ ID NO 118; 123pp; English.
XX
CC The invention comprises amino acid and coding sequences containing
CC genetic polymorphisms associated with an altered risk of developing an
CC autoimmune disease (e.g. rheumatoid arthritis). The invention further
CC comprises a method of identifying an individual that has an altered risk
CC of developing an autoimmune disease, comprising detecting a single
CC nucleotide polymorphism (SNP) in a nucleic acid of the invention. The DNA
CC and protein sequences of the invention are useful for diagnosing and
CC treating autoimmune diseases, such as: rheumatoid arthritis, type 1
CC diabetes, multiple sclerosis, systemic lupus erythematosus, inflammatory
CC bowel diseases, psoriasis, thyroiditis, celiac disease, pernicious
CC anaemia, asthma, vitiligo, glomerulonephritis, Grave's disease,
CC myocarditis, Sjogren's disease, or primary systemic vasculitis. The
CC present amino acid sequence represents a human autoimmune disease-related
CC protein of the invention. NOTE: The present sequence is not shown in the
CC specification, but has been retrieved from the WIPO website.
XX
SQ Sequence 2036 AA;
Query Match 91.2%; Score 9853; DB 8; Length 2036;
Best Local Similarity 91.5%; Pred. No. 0;
Matches 1978; Conservative 14; Mismatches 26; Indels 144; Gaps 7;
Qy 1 MTSLAQQLORLALPOSDASLSRDEVASLLFDPKEAATIDRTAFAGCTGEEELGIDP 60
Db 1 MTSLAQQLORLALPOSDASLSRDEVASLLFDPKEAATIDRTAFAGCTGEEELGIDP 60
Qy 61 SPEQFEAPLFSQAKTLERSVOTKAVNKOLDENISLFIHLSPYFLPKPAQCLEWLIHR 120
Db 61 SPEQFEAPLFSQAKTLERSVOTKAVNKOLDENISLFIHLSPYFLPKPAQCLEWLIHR 120
Qy 121 FHIHLYNQDSLACVLPYHETRFVRYIQLKINNSKRWFWLLPVKQSGVPLAKGTLIT 180
Db 121 FHIHLYNQDSLACVLPYHETRFVRYIQLKINNSKRWFWLLPVKQSGVPLAKGTLIT 180
Qy 181 HCYKOLGEMDFICSLVTKSVKFAEYPGSSAOLRVLLAFYASTIVSALVAEDVSDNIIA 240
Db 181 HCYKOLGEMDFICSLVTKSVKFAEYPGSSAOLRVLLAFYASTIVSALVAEDVSDNIIA 240
Qy 241 KLFPIYQKGLKSSLPDYRAATYMIICQISVKVTM---ENTFVNSLASQ-IITLTKIPS- 295
Db 241 KLFPIYQKSL-----YLTGRRTEQTVSLRHQNIQDQSLFDQGWKLLDSAPAE 289
Qy 296 -----LKDGLSCLIVLLQROKPSLGKPPPHLCNVDPDLTILHGISETYD 342
Db 296 TEAREBPWEKVCVTELRNGAS-----QRPFPHLNCVDPDLTILHGISETYD 334
Qy 343 VSPLLRYMLPHLVVSIHHVTGTEGMDGQIYKRLHLEAILTKISLKNLDDLHLLSLFPE 402
Db 335 VSPLLRYMLPHLVVSIHHVTGTEGMDGQIYKRLHLEAILTKISLKNLDDLHLLSLFPE 394
Qy 403 EYISYSSQEMDSNKKVSLNEQFLPLIRLESKYPRTLDVLEEHLKEIADLKQELFHQ 462
Db 395 EYISYSSQEMDSNKKVSLNEQFLPLIRLESKYPRTLDVLEEHLKEIADLKQELFHQ 454
Qy 463 FVSLSTSGKYQFLADSDTSLMLSNHPLAPVRILAMNHLKIMKTSKEGVDESFIKEAV 522
Db 455 FVSLSTSGKYQ-----EGVDESFIKEAV 478
Qy 523 LARLGDNDIDVLSAISAFEIKFHPSSSVTISNLLNLFORAEKSNKGWYEVYLKIAADI 582
Db 479 LARLGDNDIDVLSAISAFEIKFHPSSSVTISNLLNLFORAEKSNKGWYEVYLKIAADI 538
Qy 583 LIKEETILSENDQLSNQVVVCLLPFVFINNDTSESAMKIAIYLSKSGICSLHPLLRGWEE 642
Db 539 LIKEETILSENDQLSNQVVVCLLPFVFINNDTSESAMKIAIYLSKSGICSLHPLLRGWEE 598

QY 643 ALENVIKSTKPGKLGIVANQKMIETLLADNINLGDPSMLKQWBDLISVGEESFNLKQKV 702
DB |||||
QY 599 ALENVIKSTKPGKLGIVANQKMIETLLADNINLGDPSMLKM----- 639
DB |||||
QY 703 TPHVILSVLSCSSSLKETHFPFAIRVFSLLQKKIKKLSBSVITAVEIPSEWHIELMLDRG 762
DB |||||
QY 763 IPVELWAHYVEELNSTQRAVEDSVFLVSLKKFIYALKAPKSPFKGDIWNPEQKEDS 822
DB |||||
QY 655 IPVELWAHYVEELNSTQRAVEDSVFLVSLKKFIYALKAPKSPFKGDIWNPEQKEDS 714
DB |||||
QY 823 RYDLHLILGLFEMMLNGADAVHPRVLMKLFPIKVHLEDVQLFKPCSVLMTYSSLSNPLN 882
DB |||||
QY 715 RYDLHLILGLFEMMLNGADAVHPRVLMKLFPIKVHLEDVQLFKPCSVLMTYSSLSNPLN 774
DB |||||
QY 883 CSVKTVLQTOALVXGCAMLSQKTOCKHOLASTSSPVVTSLLNLNGSPKVEVRRRAIOCL 942
DB |||||
QY 775 CSVKTVLQTOALVXGCAMLSQKTOCKHOLASTSSPVVTSLLNLNGSPKVEVRRRAIOCL 834
DB |||||
QY 943 QALSGVASPFYLIIDHLISKAEBITSDAAYVIOQLATLFEELQREKKLKHQKLSLTKN 1002
DB |||||
QY 835 QALSGVASPFYLIIDHLISKAEBITSDAAYVIOQLATLFEELQREKKLKHQKLSLTKN 894
DB |||||
QY 1003 LLSCVTSCPSYIAKOLMKVLQGVNGEMVLSOLLPMABQLLEKIQKPTAVLKDEAMVHL 1062
DB |||||
QY 895 LLSCVTSCPSYIAKOLMKVLQGVNGEMVLSOLLPMABQLLEKIQKPTAVLKDEAMVHL 954
DB |||||
QY 1063 TLGKYNERSVSLNEDPKSLDIFIKAVHTTKELYAGMPTIQITALEKTIKPPFAAISDEK 1122
DB |||||
QY 955 TLGKYNERSVSLNEDPKSLDIFIKAVHTTKELYAGMPTIQITALEKTIKPPFAAISDEK 1014
DB |||||
QY 1123 VQOKLRLMFDLLVNCNKHCAQTSSVFSGISVNAEQVRIELEPPDKAKPIGTVOOKR 1182
DB |||||
QY 1015 VQOKLRLMFDLLVNCNKHCAQTSSVFSGISVNAEQVRIELEPPDKAKPIGTVOOKR 1074
DB |||||
QY 1183 QRMQOKKSQDLRSVQVSGSYQORVTLILELQHKKKLRSPOILVPTLNFLLSRCLEPLP 1242
DB |||||
QY 1075 QRMQOKKSQDLRSVQVSGSYQORVTLILELQHKKKLRSPOILVPTLNFLLSRCLEPLP 1134
DB |||||
QY 1243 QROGMEYTKQILSCLLNICOKLSPDGKIPKIDLEBEKNVELVQICIRSEMPQTHH 1302
DB |||||
QY 1135 QROGMEYTKQILSCLLNICOKLSPDGKIPKIDLEBEKNVELVQICIRSEMPQTHH 1194
DB |||||
QY 1303 HALLLLGTVAGIPDPKVLHNIMSIFTFMGANVRLDDTYSFQVINKTVQWVIPALIQSDS 1362
DB |||||
QY 1195 HALLLLGTVAGIPDPKVLHNIMSIFTFMGANVRLDDTYSFQVINKTVQWVIPALIQSDS 1254
DB |||||
QY 1363 GDSIEVSRNVEEIVVKIISVFDVALPHVPEHRRLPILVOLVDTLGAEKFLWILLILLPQ 1422
DB |||||
QY 1255 GDSIEVSRNVEEIVVKIISVFDVALPHVPEHRRLPILVOLVDTLGAEKFLWILLILLPQ 1314
DB |||||
QY 1423 VYTKTVLAAAYGEKDAILEADTEFPVSCCESVQHOIQSLMNILOYLKLPPEKEETIP 1482
DB |||||
QY 1315 VYTKTVLAAAYGEKDAILEADTEFPVSCCESVQHOIQSLMNILOYLKLPPEKEETIP 1374
DB |||||
QY 1483 KAVSFNKSQSEMLQVFNVEHTSKQLRHFKFLSVSFMSSQLSSNNFLKVVESGGPEI 1542
DB |||||
QY 1375 KAVSFNKSQSEMLQVFNVEHTSKQLRHFKFLSVSFMSSQLSSNNFLKVVESGGPEI 1434
DB |||||
QY 1543 LKGLBERLLETVLGYISAVAQSMERNADKLTVKFWRALLSKAYDILLDKVNALLPTETFTIP 1602
DB |||||
QY 1435 LKGLBERLLETVLGYISAVAQSMERNADKLTVKFWRALLSKAYDILLDKVNALLPTETFTIP 1494
DB |||||
QY 1603 VIRGLVGNPLPSVRRKALDLNKLQONISWKKTIIVTRFLKVPDLIAIVQRKKEGEE 1662
DB |||||
QY 1495 VIRGLVGNPLPSVRRKALDLNKLQONISWKKTIIVTRFLKVPDLIAIVQRKKEGEE 1554
DB |||||
QY 1663 QAINRQTALYTLKLLCKNFGAENPPFPVPLTAVKLIAPERKEKNVLGSLLCIAEVT 1722
DB |||||
QY 1555 QAINRQTALYTLKLLCKNFGAENPPFPVPLTAVKLIAPERKEKNVLGSLLCIAEVT 1614
DB |||||

QY 1723 STLEALAIQPLSLMPSLLTTMKNTSELVSSEVYLSALAAQKVVTILPHFTISPYLEGI 1782
DB |||||
QY 1615 STLEALAIQPLSLMPSLLTTMKNTSELVSSEVYLSALAAQKVVTILPHFTISPYLEGI 1674
DB |||||
QY 1783 LSQVHLEKITSEMGSASQANIRLTSKKTTLATTLAPRVLIIPAIKTTYQIEKNWNG 1842
DB |||||
QY 1675 LSQVHLEKITSEMGSASQANIRLTSKKTTLATTLAPRVLIIPAIKTTYQIEKNWNG 1734
DB |||||
QY 1843 PFMSILQEHIGMKKEELTSHOSQLTAFLEALDFRAQHSENDLEEVGKTENCIIIDCLVA 1902
DB |||||
QY 1735 PFMSILQEHIGMKKEELTSHOSQLTAFLEALDFRAQHSENDLEEVGKTENCIIIDCLVA 1794
DB |||||
QY 1903 MVVKLSEVTFRPLFKLFDWAKTEADAPKORLLITFFYNLADCIABKLGKLTFLFAGHLVKPF 1962
DB |||||
QY 1795 MVVKLSEVTFRPLFKLFDWAKTEADAPKORLLITFFYNLADCIABKLGKLTFLFAGHLVKPF 1854
DB |||||
QY 1963 ADTLQOVNISKTDEAFFDSENDPEKCCLLQLFLNCLYKIFLFDTOHFIKSKERAXALMWP 2022
DB |||||
QY 1855 ADTLQOVNISKTDEAFFDSENDPEKCCLLQLFLNCLYKIFLFDTOHFIKSKERAXALMWP 1914
DB |||||
QY 2023 LVDQLENRLGSGEEKFQERVTKHLIPCIASFVAMADDSLWKPLNYQILLKTRDSSPKVRF 2082
DB |||||
QY 1915 LVDQLENRLGSGEEKFQERVTKHLIPCIASFVAMADDSLWKPLNYQILLKTRDSSPKVRF 1974
DB |||||
QY 2083 AALITVIALAABKLENYIVLLPESIPFLAELMEDECEVEHOCQKTIQOLETVLGSPLOS 2142
DB |||||
QY 1975 AALITVIALAABKLENYIVLLPESIPFLAELMEDECEVEHOCQKTIQOLETVLGSPLOS 2034
DB |||||
QY 2143 YF 2144
DB |||||
QY 2035 YF 2036
DB |||||

RESULT 8

ABG15232
ID ABG15232 standard; protein; 1569 AA.
XX
AC ABG15232;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #15223.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US008631.
XX
PR 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR N-PSDB; AAS79419.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
PS Claim 20; SEQ ID NO 45591; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain

CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (ii). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (i) is useful in gene therapy techniques to restore normal
CC activity of (ii) or to treat disease states involving (ii). (ii) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (ii) and its binding partners are useful in medical imaging
CC of sites expressing (ii). (i) and (ii) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABC00010-ABC03077 represent novel human diagnostic
CC amino acid sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX

SQ Sequence 1569 AA;

Query Match 71.8%; Score 7756.5; DB 4; Length 1569;
Best Local Similarity 93.1%; Pred. No. 0;
Matches 1553; Conservative 1; Mismatches 1; Indels 113; Gaps 3;
QY 1 MTLAQQLQRLALPOSASLLSRDEVASLLFPDKEAATIDRTAFAGCTGLEELLGIDP 60
DB 2 MTLAQQLQRLALPOSASLLSRDEVASLLFPDKEAATIDRTAFAGCTGLEELLGIDP 61
QY 61 SFEQFEAPLPSOLAKTLERSVOTKAVNKOLDENISLFLHLSPYLLKPAQCLWLIHR 120
DB 62 SFEQFEAPLPSOLAKTLERSVOTKAVNKOLDENISLFLHLSPYLLKPAQCLWLIHR 121
QY 121 FHIHLNQDSLACVLPYHETRI FVRVIOQLKINNSKRWFWLLPVKQSGVPLAKGTLLT 180
DB 122 FHIHLNQDSLACVLPYHETRI FVRVIOQLKINNSKRWFWLLPVKQSGVPLAKGTLLT 181
QY 181 HCYKDLGFMDFICSLVTKSVKVAEPYGGSSAQLRVLLAFYASTIVSALVAADSDNIIA 240
DB 182 HCYKDLGFMDFICSLVTKSVKVAEPYGGSSAQLRVLLAFYASTIVSALVAADSDNIIA 241
QY 241 KLFPYIOKGLKSLPDYRAATYMIICQISVKVTMENTFVNSLASQIIKTLTKIPSLIKDG 300
DB 242 KLFPYIOKGLKSLPDYRAATYMIICQISVKVTMENTFVNSLASQIIKTLTKIPSLIKDG 301
QY 301 LSCLI VLLQRPESLGKK-----PFPHLCNVPDLITILHIGISETYDVSPLL 347
DB 302 LSCLI VLLQRPESLGKKYVQLILRNGASQRPFPHLCNVPDLITILHIGISETYDVSPLL 361
QY 348 RYMLPHLVYSIIHVTGTEGMDGQIYKRHLEAILTKISLKNLDHLLASLLFEEYISY 407
DB 362 HYMLPHLVYSIIHVTGTEGMDGQIYKRHLEAILTKISLKNLDHLLASLLFEEYISY 421
QY 408 SSOEEMDSNKVSLNEQFLPLRLLESKYPTLDVLEHLKEIADLKKOELPHQFVSLS 467
DB 422 SSOEEMDSNKVSLNEQFLPLRLLESKYPTLDVLEHLKEIADLKKOELPHQFVSLS 481
QY 468 TSGGKYQFLADSDTSLMLSLNHPVRLPVRILAMNHLKKIMKTSKEGYDESFIKEAVLARLG 527
DB 482 TSGGKYQ-----EGVDESFIKEAVLARLG 505
QY 528 DDNIDVLSAISAFETPKHFSPSEVTISNLLNLFORAEISKNGEWTEVLKIAADILIKEE 587
DB 506 DDNIDVLSAISAFETPKHFSPSEVTISNLLNLFORAEISKNGEWTEVLKIAADILIKEE 565
QY 588 ILSENDQLSNQVVVCLLPFVWINDDTSAEMKIAIYLSKSGICSLHPLILRGWEALENV 647
DB 566 ILSENDQLSNQVVVCLLPFVWINDDTSAEMKIAIYLSKSGICSLHPLILRGWEALENV 625
QY 648 IKSTKPKGLIGVANQKMIELADNINILGDPSSMLKNVEDLISVGEBSFNLKQKVFHVI 707
DB 626 IKSTKPKGLIGVANQKMIELADNINILGDPSSMLKN----- 661

QY 708 LSVLVSCSSSLKETHFPFAIRVFSLLQKKIKKLSVITAVEIPSEWHIELMLDRGIPVEL 767
DB 662 -----BIPSEWHIELMLDRGIPVEL 681
QY 768 WAHYVEELNSTORVAVEDSVFLVFLSKKFIYALKAPKSFPPKGDIIWNPPEQLKEDSRDYLH 827
DB 682 WAHYVEELNSTORVAVEDSVFLVFLSKKFIYALKAPKSFPPKGDIIWNPPEQLKEDSRDYLH 741
QY 828 LLIGLFEMMLNGADAVHPRVLMKLFIKVHLEBDVFLQFKFCSVLWITYGSSLNPLNCSVK 887
DB 742 LLIGLFEMMLNGADAVHPRVLMKLFIKVHLEBDVFLQFKFCSVLWITYGSSLNPLNCSVK 801
QY 888 VLQTOALYVGCAMLSQKTOCKHOLASISSPVVTSLLNLGSPVKEVRRAAIQCLOALSG 947
DB 802 VLQTOALYVGCAMLSQKTOCKHOLASISSPVVTSLLNLGSPVKEVRRAAIQCLOALSG 861
QY 948 VASPFYLIIDHLISKAEBITSDAAYVIODLATLFEELQREKKLSHQKLSSETLKNLLSCV 1007
DB 862 VASPFYLIIDHLISKAEBITSDAAYVIODLATLFEELQREKKLSHQKLSSETLKNLLSCV 921
QY 1008 YSCPSYIAKOLMKVLQGVNGEMVLSQLLPMAEQLLEKIQKEPTAVLKDEAMVHLTLGKY 1067
DB 922 YSCPSYIAKOLMKVLQGVNGEMVLSQLLPMAEQLLEKIQKEPTAVLKDEAMVHLTLGKY 981
QY 1068 NEFSVSLINEDPKSLDIFIKAHVHTTKELYAGWPTTIQITALEKIKTPPFAAISDEKVOOKL 1127
DB 982 NEFSVSLINEDPKSLDIFIKAHVHTTKELYAGWPTTIQITALEKIKTPPFAAISDEKVOOKL 1041
QY 1128 LRMLFDLLVNCNKNSHCAQTVSVFKGISVNAEQVRIELEPPDKAKPLGTVQQRKQMQ 1187
DB 1042 LRMLFDLLVNCNKNSHCAQTVSVFKGISVNAEQVRIELEPPDKAKPLGTVQQRKQMQ 1101
QY 1188 KKSQLESVQEVGGSYWQRTVLIILELQHKKLRSPQILVPTLFNLLSRCLEPLPQEQGN 1247
DB 1102 KKSQLESVQEVGGSYWQRTVLIILELQHKKLRSPQILVPTLFNLLSRCLEPLPQEQGN 1161
QY 1248 MEYTKQLILSCLINICQKLSPDGGKI PKDILDEEFNVELIVQIRLSEMPQTHHALL 1307
DB 1162 MEYTKQLILSCLINICQKLSPDGGKI PKDILDEEFNVELIVQIRLSEMPQTHHALL 1221
QY 1308 LGTVAGIIPDKVLHNMISIFTFMGANVRLDDTYSFQVINKTVKVIIPALIQSDSGDSIE 1367
DB 1222 LGTVAGIIPDKVLHNMISIFTFMGANVRLDDTYSFQVINKTVKVIIPALIQSDSGDSIE 1281
QY 1368 VSRNVEEIVVKIISVFDALPHVPEHRPLPIVLQVLDTLGAKFLWILLILLFEQVYTKT 1427
DB 1282 VSRNVEEIVVKIISVFDALPHVPEHRPLPIVLQVLDTLGAKFLWILLILLFEQVYTKT 1341
QY 1428 VLAAYGEKDALTEADTFWFSVCCFVSVOHQIQSLMNLQVLLKLPPEKEETIPKAVSF 1487
DB 1342 VLAAYGEKDALTEADTFWFSVCCFVSVOHQIQSLMNLQVLLKLPPEKEETIPKAVSF 1401
QY 1488 NKSEQSEEMLVNFVETHTSKQLRHFKFLSVFMSQLSSNNFLKVVESGGPPEILKGL 1547
DB 1402 NKSEQSEEMLVNFVETHTSKQLRHFKFLSVFMSQLSSNNFLKVVESGGPPEILKGL 1461
QY 1548 ERLLTETVLGYISAVAQSMERNADKLTWKFWRALSKAYDLDKVNALLPTETFIPIVIRGL 1607
DB 1462 ERLLTETVLGYISAVAQSMERNADKLTWKFWRALSKAYDLDKVNALLPTETFIPIVIRGL 1521
QY 1608 VGNPLPSVRRKALDILNNKLQNIISWKKTIIVTRFLKLVDPDLIAIVQRK 1655
DB 1522 VGNPLPSVRRKALDILNNKLQNIISWKKTIIVTRFLKLVDPDLIAIVQRK 1569

RESULT 9

ADQ66400

ID ADQ66400 standard; protein; 1229 AA.

XX ADQ66400;

XX ADQ66400;

DT 07-OCT-2004 (first entry)

XX Novel human protein sequence #1373.
 DE osteopathic; neuroprotective; nootropic; antiparkinsonian; cytostatic;
 XX gene therapy; diagnostic marker; morbid state; osteoporosis;
 KW neurological disease; Alzheimer's disease; Parkinson's disease; dementia;
 KW cancer.
 XX Homo sapiens.
 OS EP1440981-A2.
 XX PN 28-JUL-2004.
 XX PD 21-JAN-2004; 2004EP-00001196.
 XX PF 21-JAN-2003; 2003JP-00102206.
 XX PR 09-MAY-2003; 2003JP-00131392.
 XX PS (REAS-) RES ASSOC BIOTECHNOLOGY.
 XX PA Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
 XX PI Yamamoto J, Isono Y, Nagai K, Irie R;
 XX DR WPI: 2004-535376/52.
 XX DR N-PSDB; ADQ64212.
 XX PT Novel 2495 cDNA, useful for treating osteoporosis, neurological diseases,
 PT Alzheimer's diseases, Parkinson's diseases, dementia and various cancers.
 XX PS Claim 1; SEQ ID NO 3561; 2449pp; English.
 XX CC The invention relates to 2495 novel polynucleotides (1) and their encoded
 CC polypeptides, sequences hybridizing to these nucleotides, sequences
 CC encoding partial polypeptides and sequences having 70% or 90% identity to
 CC the nucleotide and protein sequences. The nucleotides and polypeptides
 CC are useful as diagnostic markers or therapeutic target for the diseases
 CC or morbid states. They are also useful for treating osteoporosis,
 CC neurological diseases, Alzheimer's diseases, Parkinson's diseases,
 CC dementia and various cancers. This sequence corresponds to a protein
 CC sequence of the invention.
 XX SQ Sequence 1229 AA;
 Query Match 56.7%; Score 6124.5; DB 8; Length 1229;
 Best Local Similarity 93.4%; Pred. No. 0;
 Matches 1223; Conservative 0; Mismatches 6; Indels 81; Gaps 1;
 QY 835 MWLNGADAVHFRVLMKLFKIHLEDFVQLFKFCSVLWYTGSSLNPLNCSVKTVLQDAL 894
 DB 1 MWLNGADAVHFRVLMKLFKIHLEDFVQLFKFCSVLWYTGSSLNPLNCSVKTVLQDAL 60
 QY 895 YVGCAMLSQKTOCKHQLASISSPVVTSLLNLGSPVKEVRRAAIQCLQALSGVASPFYL 954
 DB 61 YVGCAMLSQKTOCKHQLASISSPVVTSLLNLGSPVKEVRRAAIQCLQALSGVASPFYL 120
 QY 955 IIDHLISKAEITSDAAVYIQDALTFEELQREKKLKHOKLSETLKNLLSCVSCPSYI 1014
 DB 121 IIDHLISKAEITSDAAVYIQDALTFEELQREKKLKHOKLSETLKNLLSCVSCPSYI 180
 QY 1015 AKDLMKVLQVNGEMVLSQLPMAEQLEKIQEPTAVLKDEAMVHLHTLKGKNEFVSVL 1074
 DB 181 AKDLMKVLQVNGE----- 194
 QY 1075 LNEPDKSLDIFIKAVHTTKELVAGMPTIQTALKEKITKPPFAAISDEKVOQKLRLMFLDL 1134
 DB 195 -----ITKPPFAAISDEKVOQKLRLMFLDL 219
 QY 1135 LVNCKNSHCAQTSSVFKGISVNAEQVRIEPPDKAKPLGTVOQKRQKMQKKSQDLE 1194
 DB 220 LVNCKNSHCAQTSSVFKGISVNAEQVRIEPPDKAKPLGTVOQKRQKMQKKSQDLE 279
 QY 1195 SVQEVGGSYWQWRTVLIILELLOHKKLRSPQILVPTLFNLLSRCLEPLPQEQGNMEYTKQL 1254

Db 280 SVQEVGGSYWQWRTVLIILELLOHKKLRSPQILVPTLFNLLSRCLEPLPQEQGNMEYTKQL 339
 QY 1255 ILSCLLNTICOKLSPDGGKIPKDLIDEEKFNVELIVQCIIRLSSEMPQTHHALLLGTVAGI 1314
 Db 340 ILSCLLNTICOKLSPDGGKIPKDLIDEEKFNVELIVQCIIRLSSEMPQTHHALLLGTVAGI 399
 QY 1315 FPDKVLHNMISITFTMGANVMRLDDTYSFQVINKTVKMWIPALIQSDSGDSIEVSRNVEE 1374
 Db 400 FPDKVLHNMISITFTMGANVMRLDDTYSFQVINKTVKMWIPALIQSDSGDSIEVSRNVEE 459
 QY 1375 IIVKLIISVFDALPHVPEHRRLPILVOLVDTLGAKEFLWILLILFEQVVTYKVLAAAYG 1434
 Db 460 IIVKLIISVFDALPHVPEHRRLPILVOLVDTLGAKEFLWILLILFEQVVTYKVLAAAYG 519
 QY 1435 EKDAILEADTEFWFVSCCEFSVQHQIQSLMNLQVLLKLPKEKEETIPKAVSFNKSSEOB 1494
 Db 520 EKDAILEADTEFWFVSCCEFSVQHQIQSLMNLQVLLKLPKEKEETIPKAVSFNKSSEOB 579
 QY 1495 EMLQVFNVEHTTSKOLRHFKFLSVFSFMSQLSSNNFLKKVSVESGGEPEILKGLERLLETV 1554
 Db 580 EMLQVFNVEHTTSKOLRHFKFLSVFSFMSQLSSNNFLKKVSVESGGEPEILKGLERLLETV 639
 QY 1555 LGYISAVAQSMERNADKLTVKFWRALISKAYDLDKVNALLPTETFIPVIRGLVGNPLPS 1614
 Db 640 LGYISAVAQSMERNADKLTVKFWRALISKAYDLDKVNALLPTETFIPVIRGLVGNPLPS 699
 QY 1615 VRRKALDLNNKLQONISWKKTIVTRFLKLPVLDLAIIVQRKKKEGEEQINQRTALYTL 1674
 Db 700 VRRKALDLNNKLQONISWKKTIVTRFLKLPVLDLAIIVQRKKKEGEEQINQRTALYTL 759
 QY 1675 KLLCKNFGAENPDPPVPLXTAVKLIAPERKEEKNVLSGALLCIAEVTSTLEALAIPOLP 1734
 Db 760 KLLCKNFGAENPDPPVPLXTAVKLIAPERKEEKNVLSGALLCIAEVTSTLEALAIPOLP 819
 QY 1735 SIMPSLLTMMKNTSELVSSEVYLSALAALQKVETLPHFTSPYLEGILSQVHLEKITS 1794
 Db 820 SIMPELLTMMKNTSELVSSEVYLSALAALQKVETLPHFTSPYLEGILSQVHLEKITS 879
 QY 1795 EMGASQANIRLTSLKKTATTAPRVLLPAIKTKYKQIEKNWKHMGPPMSILOEHIGX 1854
 Db 880 EMGASQANIRLTSLKKTATTAPRVLLPAIKTKYKQIEKNWKHMGPPMSILOEHIGX 939
 QY 1855 MKKEELTSHQSLTAFFLEALDFRAQSENDELEVGTENCIIIDCLVAMVVKLSVTRFP 1914
 Db 940 MKKEELTSHQSLTAFFLEALDFRAQSENDELEVGTENCIIIDCLVAMVVKLSVTRFP 999
 QY 1915 LFFKLFDWAKTEDAPKORLLTFYNLADCIABKLGKLFPLPAGHLVKPFDATLXQVNIKT 1974
 Db 1000 LFFKLFDWAKTEDAPKORLLTFYNLADCIABKLGKLFPLPAGHLVKPFDATLXQVNIKT 1059
 QY 1975 DEAFDSENDEPEKCLLLOFNLCLYKIFLFDTOHFISKERAXALMPLVDQLENRLGE 2034
 Db 1060 DEAFDSENDEPEKCLLLOFNLCLYKIFLFDTOHFISKERAXALMPLVDQLENRLGE 1119
 QY 2035 EKQERVTKHLIPCIAQFSVAMADDSLMKPLNYQILKTRDSSPKVRFAALITVLALAEK 2094
 Db 1120 EKQERVTKHLIPCIAQFSVAMADDSLMKPLNYQILKTRDSSPKVRFAALITVLALAEK 1179
 QY 2095 LKENYIVLLPESIPFLAELMEDECEEVHQCQKTIQOLETVLGEPLOSIF 2144
 Db 1180 LKENYIVLLPESIPFLAELMEDECEEVHQCQKTIQOLETVLGEPLOSIF 1229
 RESULT 10
 ADE08012
 ID ADE08012 standard; protein; 1149 AA.
 XX
 AC ADE08012;
 XX
 DT 29-JAN-2004 (first entry)
 XX
 DE Novel protein (useful for identifying genetic disorders) #167.

XX novel gene; novel protein; tissue marker; molecular weight marker;
KW chromosome marker; genetic disorder.
XX Unidentified.
XX WO2003054152-A2.
XX 03-JUL-2003.
XX 10-DEC-2002; 2002WO-US039555.
XX 10-DEC-2001; 2001US-0339739P.
XX 11-DEC-2001; 2001US-0339453P.
XX 14-MAR-2002; 2002US-0365091P.
XX 14-MAR-2002; 2002US-0365384P.
XX 12-APR-2002; 2002US-0372381P.
XX 12-APR-2002; 2002US-0372615P.
XX 22-APR-2002; 2002US-00128558.
XX 24-APR-2002; 2002US-0376045P.
XX (HYSE-) HYSEQ INC.
XX Tang YT, Asundi V, Goodrich RW, Ren F, Zhang J, Zhao QA, Wang J;
PI Ghosh M, Xue AJ, Wehrman T, Weng G, Zhou P, Drmanac RT, Wang Z;
PI Ma Y, Wang D, Chen R, Xu C, Boyle BJ;
XX WPI; 2003-569235/53.
XX N-PSDB; ADE07101.
XX New polynucleotides, useful for expressing recombinant proteins for
PT analysis, characterization or therapeutic use, or as markers for tissues
PT in which the corresponding protein is preferentially expressed.
XX Claim 20; SEQ ID NO 1078; 1177pp; English.
XX The invention comprises the amino acid and coding sequences of novel
CC proteins. The DNA and protein sequences of the invention are useful as:
CC markers for tissues in which the corresponding protein is preferentially
CC expressed; as molecular weight markers on gels; as chromosome markers or
CC tags; to identify chromosomes or to map related gene positions; and to
CC compare with endogenous DNA sequences in patients to identify potential
CC genetic disorders. The present amino acid sequence represents a protein
CC of the invention.
XX
XX Sequence 1149 AA;
Query Match 46.5%; Score 5026.5; DB 7; Length 1149;
Best Local Similarity 52.4%; Pred. No. 0;
Matches 1123; Conservative 7; Mismatches 19; Indels 995; Gaps 14;
QY 1 MTSLAQQLQRLALPQSDASILSRDEVASLLFDPKAAATIDRTAFAGTCTGLEELGIDP 60
DB 1 MTSLAQQLQRLALPQSDASILSRDEVASLLFDPKAAATIDRTAFAGTCTGLEELGIDP 60
QY 61 SFEQFEAPLFSQAKTLERSVQTKAVNKQLDENISLFLHLSPYFLPKAQCLEWLIHR 120
DB 61 SFEQFEAPLFSQAKTLERSVQTKA----- 85
QY 121 FTHILYNQDSLIACVLPYHETRFVVRVIOQLKINNSKRWFWLLPVKQSGVPLAKGTLLT 180
DB 86 -----QSGVPLAKGTLLT 98
QY 181 HCYKDLGFMDFICSLVTKSVKVFAPYGGSAQRLVLLAFYASTIVSALVAEDVSNIIA 240
DB 99 HCYKDLGFMDFICSLVTKSVKVFAPYGGSAQRLVLLAFYASTIVSALVAEDVSNIIA 158
QY 241 KLPFYQKGLKSSLPDYRAATYMIQISVKVTWENTFVNSLASQIKTLTKTIPSLIKOG 300
DB 159 KLPFYQK-----IITKTKIPSLIKOG 181
QY 301 LSCILVLLQKQPESLGKPPHLCNVPDLITILHGISETYDVSPLRLYMLPHLWVSIH 360
DB 182 LSCILVLLQKQPESLGK----- 200
QY 361 HVTGEETEGMDGQIYKRHLRAILTKISLKNLHLLASLLFEEYISYSSQEMDSNKKVSL 420
DB 201 ----- 200
QY 421 LNEQFLPLIRLLESKYPRTLDDVVLEBHLKETADLUKQELFHQFVLSLTSGGKYOFLADSD 480
DB 201 ----- 200
QY 481 TSLMSLNHPLAPVRILAMNHLKKIMKTSKEGVDESFIKEAVLARLGDNDIDVLSAISA 540
DB 201 ----- 200
QY 541 FEIPEHFSSEVTTISNLLNLFQRAELSKNGEWYEVLTIAADILIKEEILSENDQLSNOVV 600
DB 201 -----YEVLKIAADILLIKEEILSENDQLSNOVV 228
QY 601 VCLLPFVIVNDDTESAEMKIAIYLSKSGICSLHPLRGWEALENVIKSTKPGKLI GVA 660
DB 229 VCLLPFVIVNDDTESAEMKIAIYLSKSGICSLHPLRGWEE----- 270
QY 661 NQMIELLADNINLGDPSMLKMVEDLISVGBEESFNLKQKTFHVILSVLSCSSSLKE 720
DB 271 ----- 270
QY 721 THFFPAIRVFSLLQKKIKKLESVITAVEIPSEWHIELMLDRGIPVELWAHYVEELNSTQR 780
DB 271 ----- 270
QY 781 VAVEDSVFLVSLKFIYALKAPKSPFGKDIWNPEQLKEDSRDYLLHLLIGLFEMWLNKA 840
DB 271 -----GDIWNPEQLKEDSRDYLLHLLIGLFEMWLNKA 302
QY 841 DAVHFRVLMKLFVKVHLEDVDFQLFKFCVLTWYTGSSLSNPLNCSVKTVLQOALYVGCAM 900
DB 303 DAVHFRVLMKLFK----- 316
QY 901 LSSQKTOCKHQLASISSPVVTTSLNLGSPVKEVRAAIQCLQALSGVASPFYLIIDHLI 960
DB 317 ----- 316
QY 961 SKAEBITSDAAAYVIQDLATLFELOREKKLKHQKSETLKNLSCVYSCPSYIAKDLMK 1020
DB 317 ----- 316
QY 1021 VLQGVNGEMVLSQLLPMAEQLEKIQKEPTAVLKDEAMVHLTLGKYNFYSVLINEDPK 1080
DB 317 -----MVLSQLPMAEQLEKIQKEPTAVLKDEAMVHLTLGKYNFYSVLINEDPK 368
QY 1081 SLDIFIKAVHTTKELYAGMPTTQITALEKIKTPEFFAAISDEKVOQKLLRMLFDLLVNCN 1140
DB 369 SLDIFIKAVHTTKELYAGMPTTQITALEKILIRY-----ISD----- 405
QY 1141 SHCAQTSSVSVFKGISVNAEQVRIELEPPDKAPLGTVOQKRRQKMQKKSQDLESVOEVG 1200
DB 406 ----- 405
QY 1201 GSYQWRVTLILELLOHKKKLRSPQIIVPTLFNLLSRCLEPLPQEQGNMEYTKQILISCLL 1260
DB 406 -----PAL-----GNNR-----CLL 415
QY 1261 NTCOKLSPGGKTPKDIIDLEEFKNVELIVQCIRLSMPOTHHALLLGTVAGIFDPKVL 1320
DB 416 NTCOKLSPGGKTPKDIIDLEEFKNVELIVQCIRLSMPOTHHALLLGTVAGIFDPKVL 475
QY 1321 HNIMSIFFWGANVMELEDDTYSFQVINKTVKVI PALIQSDSGDSIEVSRNVEEIVVKII 1380
DB 476 HNIMSIFFWGANVMELEDDTYSFQVINKTVKVI PALIQSDSGDSIEVSRNVEEIVVKII 535
QY 1381 SVFVDALPHVPEHRRPLPILVOLVDTLGAEKFLWILLILPEQVYVTKVLAAYAGEKDAIL 1440
DB 536 SVFVDALPHVPEHRRPLPILVOLVDTLGAEKFLWILLILPEQVYVTKVLAAYAGEKDAIL 595

CC	The invention relates to purified proteins and genes encoding them, that are involved in T cell activation (1) and has an amino acid deletion, substitution or addition in the amino acid sequences. The methods and compositions of the present invention are useful for the diagnosis, prevention and/or treatment of autoimmune disease (rheumatoid arthritis, asthma, multiple sclerosis and diabetes), allergic disease, infectious disease, AIDS, and acute or chronic rejection at organ transplant or bone marrow transplant. This sequence corresponds to a protein involved in T cell activation.
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	
XX	

```

Db      841 VLALAEKLNKENVYLLPESIPFLAELMEDECEVEHQCKTIQOLETVLGEPLQSYF 897
|||||
RESULT 13
ADQ96152
ID      ADQ96152 standard; protein; 897 AA.
XX
AC      ADQ96152;
XX
DT      07-OCT-2004 (first entry)
XX
DE      T cell activation associated protein #165.
XX
KW      antiallergic; antiarthritic; antiasthmatic; antidiabetic; anti-HIV;
KW      antimicrobial; antirheumatic; immunosuppressive; neuroprotective;
KW      gene therapy; T cell activation; diagnosis; autoimmune disease;
KW      rheumatoid arthritis; asthma; multiple sclerosis; diabetes;
KW      allergic disease; infectious disease; AIDS; chronic rejection; organ;
KW      bone-marrow transplant.
XX
OS      Homo sapiens.
XX
PN      WO2004058805-A2.
XX
PD      15-JUL-2004.
XX
PF      25-DEC-2003; 2003WO-JP016715.
XX
PR      26-DEC-2002; 2002JP-00376365.
PR      27-DEC-2002; 2002US-0436473P.
PR      25-APR-2003; 2003JP-00122113.
PR      28-APR-2003; 2003US-0465792P.
PR      21-OCT-2003; 2003JP-00360559.
PR      22-OCT-2003; 2003US-0512846P.
XX
PA      (ASAH-) ASahi Kasei Pharma Corp.
XX
PI      Matsuda A, Yoneta S;
XX
WPI: 2004-593134/57.
DR      N-PSDB; ADQ96151.
XX
XX      New purified protein involved in T cell activation, useful for
PT      diagnosing, preventing and/or treating acquired immunodeficiency
PT      syndrome, autoimmune (e.g. rheumatoid arthritis, and diabetes), allergic
PT      and infectious diseases.
XX
PS      Claim 1; SEQ ID NO 330; 2828pp; English.
XX
CC      The invention relates to purified proteins and genes encoding them, that
CC      are involved in T cell activation (I) and has an amino acid deletion,
CC      substitution or addition in the amino acid sequences. The methods and
CC      compositions of the present invention are useful for the diagnosis,
CC      prevention and/or treatment of autoimmune disease (rheumatoid arthritis,
CC      asthma, multiple sclerosis and diabetes), allergic disease, infectious
CC      disease, AIDS, and acute or chronic rejection at organ transplant or bone
CC      marrow transplant. This sequence corresponds to a protein involved in T
CC      cell activation.
XX
SQ      Sequence 897 AA;

Query Match      41.6%; Score 4494; DB 8; Length 897;
Best Local Similarity 99.2%; Pred. No. 1.7e-317;
Matches 890; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY      1248 MEYTKQLILSCLLNLCOKLSPDGGIKPKDILDEEFNVELIVQCIRLSEMPQTHHALLL 1307
DB      1 MEYTKQLILSCLLNLCOKLSPDGGIKPKDILDEEFNVELIVQCIRLSEMPQTHHALLL 60
QY      1308 LGTVAGIPDPKVLHNMISIFTFMGANVRLDDTYSFQVINKTVKMWIPALIQSDSGDSIE 1367
DB      61 LGTVAGIPDPKVLHNMISIFTFMGANVRLDDTYSFQVINKTVKMWIPALIQSDSGDSIE 120

```

```

QY      1368 VSRNVEEIVVKIISVFVDALPHVPEHRRLPILVQLVDTLGAEKFLWILLILLPEQVYTKT 1427
DB      121 VSRNVEEIVVKIISVFVDALPHVPEHRRLPILVQLVDTLGAEKFLWILLILLPEQVYTKT 180
QY      1428 VLAAYGEKDAILEADTEFWFSCVCEFSVQHQIQSLMNIQYLLKLPESKEETIPKAVSF 1487
DB      181 VLAAYGEKDAILEADTEFWFSCVCEFSVQHQIQSLMNIQYLLKLPESKEETIPKAVSF 240
QY      1488 NKSESOEMLQVFNVEHTSKLRHFKPLSVFMSQLSSNNFLKVVSGGGPEILLKGLE 1547
DB      241 NKSESOEMLQVFNVEHTSKLRHFKPLSVFMSQLSSNNFLKVVSGGGPEILLKGLE 300
QY      1548 ERLLETVLGYISAVAQSMERNADKLTVKFWRALLSKAYDILLDKVNALLTETPIPIVIRGL 1607
DB      301 ERLLETVLGYISAVAQSMERNADKLTVKFWRALLSKAYDILLDKVNALLTETPIPIVIRGL 360
QY      1608 VGNPLPSVRKALDILNNKLOQNISWKKITVTRFKLVPDILAIIVQRKKKEBEEQAINR 1667
DB      361 VGNPLPSVRKALDILNNKLOQNISWKKITVTRFKLVPDILAIIVQRKKKEBEEQAINR 420
QY      1668 QTALYTLKLLCKNFGAENPDPFVPLXTAVKLIAPERKEKNVLGSALLCIAEVTSTLEA 1727
DB      421 QTALYTLKLLCKNFGAENPDPFVPLXTAVKLIAPERKEKNVLGSALLCIAEVTSTLEA 480
QY      1728 LAIPQLPSLMPSLLTMTKNTSELVSEVYLLSALAALQKVETLPHFISPYLEGILSQVI 1787
DB      481 LAIPQLPSLMPSLLTMTKNTSELVSEVYLLSALAALQKVETLPHFISPYLEGILSQVI 540
QY      1788 HLEKITSEMGASQANIRLTSKKTLATTLAPRVLLPAIKKTYKQIEKNWKNHMGPFMSI 1847
DB      541 HLEKITSEMGASQANIRLTSKKTLATTLAPRVLLPAIKKTYKQIEKNWKNHMGPFMSI 600
QY      1848 LOEHIGXMKKEELTSHQSOLTAFFLEALDFRAQHSNDLEEVGKTENCIIICLVAMVVKL 1907
DB      601 LOEHIGXMKKEELTSHQSOLTAFFLEALDFRAQHSNDLEEVGKTENCIIICLVAMVVKL 660
QY      1908 SEVTFRPLFFKLFDWAKTEDAPKDRLLTFYNLADCIAEKLGLFTLFAGHLVKPFADTLX 1967
DB      661 SEVTFRPLFFKLFDWAKTEDAPKDRLLTFYNLADCIAEKLGLFTLFAGHLVKPFADTLX 720
QY      1968 QVNISKTDPAFPDSENDPEKCLLLQFILNCLYKIFLFDTHQHFISKERAXALMPLVDQL 2027
DB      721 QVNISKTDPAFPDSENDPEKCLLLQFILNCLYKIFLFDTHQHFISKERAXALMPLVDQL 780
QY      2028 ENRLGGEKGFQBRVTKHLIPCTAQSVMADDSLWKPLNYQILLKTRDSSPKVRFAALIT 2087
DB      781 VNRLGGEKGFQBRVTKHLIPCTAQSVMADDSLWKPLNYQILLKTRDSSPKVRFAALIT 840
QY      2088 VLALAEKLNKENVYLLPESIPFLAELMEDECEVEHQCKTIQOLETVLGEPLQSYF 2144
DB      841 VLALAEKLNKENVYLLPESIPFLAELMEDECEVEHQCKTIQOLETVLGEPLQSYF 897

RESULT 14
ADQ04615
ID      ADQ04615 standard; protein; 734 AA.
XX
AC      ADQ04615;
XX
DT      20-MAY-2004 (first entry)
XX
DE      Human protein of the invention SEQ ID NO:3300.
XX
KW      human; gene therapy; diagnostic marker; pharmaceutical.
XX
OS      Homo sapiens.
XX
PN      EP1347046-A1.
XX
PD      24-SEP-2003.
XX
PF      12-APR-2002; 2002EP-00008400.

```

XX 22-MAR-2002; 2002JP-00137785.
XX (REAS-) RES ASSOC BIOTECHNOLOGY.
XX Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
PI Yanamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
PI Seki N, Yoshikawa T, Otsuka M, Nagahashi K, Masuho Y;
XX WPI; 2003-723558/69.
DR N-PSDB; ADM02172.
XX
PT New polynucleotides and polypeptides are useful in gene therapy, for
PT developing a diagnostic marker or medicines for regulating their
PT expression and activity, or as a target of gene therapy.
XX
PS Claim 1; SEQ ID NO 3300; 305pp; English.
XX
XX The invention relates to a novel human polynucleotide and the encoded
CC polypeptide. A polynucleotide of the invention may have a use in gene
CC therapy. An oligonucleotide of the invention ADM06202-ADM06773 is useful
CC as a primer for synthesizing the polynucleotide or as a probe for
CC detecting the polynucleotide. The polynucleotides ADM01316-ADM03758 are
CC useful in gene therapy, for developing a diagnostic marker or medicines
CC for regulating their expression and activity, or as a target of gene
CC therapy. The proteins ADM03759-ADM06201 encoded by the polynucleotides
CC are useful as pharmaceutical agents. The present sequence represents a
CC protein sequence of the invention.
XX
SQ Sequence 734 AA;
Query Match 34.2%; Score 3694; DB 7; Length 734;
Best Local Similarity 100.0%; Pred. No. 1.7e-259;
Matches 734; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MTSLAQQLQRLALPQSDASLLSRDEVASLLFPDKAAATIDRTAFAGTGTGEEELGIDP 60
DB 1 MTSLAQQLQRLALPQSDASLLSRDEVASLLFPDKAAATIDRTAFAGTGTGEEELGIDP 60
QY 61 SPEQEPAPLFSOLAKTLERSVQTKAVNKOLDENISLFLHLSPYFLPKPAQKCLEWLIHR 120
DB 61 SPEQEPAPLFSOLAKTLERSVQTKAVNKOLDENISLFLHLSPYFLPKPAQKCLEWLIHR 120
QY 121 FTHILYNQDSLACVLPYHETRFVRIQLLKINNSKRWFWLLPVKQSGVPLAKGTLLT 180
DB 121 FTHILYNQDSLACVLPYHETRFVRIQLLKINNSKRWFWLLPVKQSGVPLAKGTLLT 180
QY 181 HCYKDLGFMDFICSLVTKSVKPAEYPGSSAQLRVLLAFYASTIVSALVAEEDVSNIIA 240
DB 181 HCYKDLGFMDFICSLVTKSVKPAEYPGSSAQLRVLLAFYASTIVSALVAEEDVSNIIA 240
QY 241 KLPPYIQKGLKSLPYRAATYMIICQISVKVTMENTFVNSLASQIKTLTKIPSLIKDG 300
DB 241 KLPPYIQKGLKSLPYRAATYMIICQISVKVTMENTFVNSLASQIKTLTKIPSLIKDG 300
QY 301 LSLCLIVLQKQKPESLGKPPFPHLCNVPLDITLHIGSETYDVSPLLRYMLPHLVVSIIT 360
DB 301 LSLCLIVLQKQKPESLGKPPFPHLCNVPLDITLHIGSETYDVSPLLRYMLPHLVVSIIT 360
QY 361 HVTGEETEGMDGQIYKRLHLEAITLTKISLKNLDHLASLLFEBYIYSYQOEEDMSKNVSL 420
DB 361 HVTGEETEGMDGQIYKRLHLEAITLTKISLKNLDHLASLLFEBYIYSYQOEEDMSKNVSL 420
QY 421 LNEQFLPLIRLESKYPTLDVLEHLKEIADLKQELPHQFVSLSTSGKQYFLADSD 480
DB 421 LNEQFLPLIRLESKYPTLDVLEHLKEIADLKQELPHQFVSLSTSGKQYFLADSD 480
QY 481 TSLMLSLNHPPLAPVIRILAMNHLKKIMKTSKEGVDSFIKEAVLARLGDDNIDVLSAISA 540
DB 481 TSLMLSLNHPPLAPVIRILAMNHLKKIMKTSKEGVDSFIKEAVLARLGDDNIDVLSAISA 540
QY 541 FEIFKEHFSSEVTISNLLNLFQRAELSKNGEWYEVLFKIAADILIKEILSENDOLSNOVV 600

DB 541 FEIFKEHFSSEVTISNLLNLFQRAELSKNGEWYEVLFKIAADILIKEILSENDOLSNOVV 600
QY 601 VCLLPFFVINDDTESAEMKIAIYLSKSGICSLHPLLRGWEEALENVIKSTPKGLIGVA 660
DB 601 VCLLPFFVINDDTESAEMKIAIYLSKSGICSLHPLLRGWEEALENVIKSTPKGLIGVA 660
QY 661 NQKMIELLADNINIGDPSSMLKMWEDLISVGEESFNLKQKVTFFHVLVSVLSCSSLUKE 720
DB 661 NQKMIELLADNINIGDPSSMLKMWEDLISVGEESFNLKQKVTFFHVLVSVLSCSSLUKE 720
QY 721 THPPFAIRVFSLLQ 734
DB 721 THPPFAIRVFSLLQ 734
RESULT 15
AAWS4099
ID AAWS4099 standard; protein; 515 AA.
XX
AC AAWS4099;
XX
DT 28-SEP-1998 (first entry)
XX
DE Homo sapiens BAP28 sequence.
XX
KW BARD1; ring protein; BRCA1; breast cancer; risk; diagnosis.
XX
OS Homo sapiens.
XX
PN WO9812327-A2.
XX
PD 26-MAR-1998.
XX
PF 19-SEP-1997; 97WO-US016842.
XX
PR 20-SEP-1996; 96US-0025296P.
PR 03-APR-1997; 97US-0042611P.
PR 04-APR-1997; 97US-0042985P.
XX
PA (TEXA) UNIV TEXAS SYSTEM.
XX
PI Bowcock AM, Baer R;
XX
DR WPI; 1998-230317/20.
DR N-PSDB; AAV24135.
XX
PT DNA sequence encoding BARD1, B123, BE2, BE14, BE31 or BE445 - which as
PT breast cancer antigen, BRCA1, binding proteins are useful to identify
PT patient having or at risk of developing cancer.
XX
PS Disclosure; Page 287-288; 348pp; English.
XX
XX The sequence is that of a protein which can be used in the preparation of
CC the recombinant breast cancer antigen, BRCA1, binding proteins BARD1,
CC B123, BE2, BE14, BE31 or BE445, or a composition for the detection of a
CC BARD1, B123, BE2, BE14, BE31 or BE445 nucleic acid sequence, specifically
CC a wild type BARD1 composition for the detection or purification of BRCA1,
CC useful to identify a patient having, or at risk of developing cancer.
CC BARD1 can be used in the preparation of an anti-BARD1 antibody, and in
CC the detection and purification of a BRCA1 protein. BARD1, B123, BE2,
CC BE14, BE31 or BE445 can be used in the identification of a binding protein
CC agonist or antagonist that alters the binding of BARD1, B123, BE2, BE14,
CC BE31 or BE445 to BRCA1 or the biological activity of the BRCA1-BARD1,
CC detect BARD1, B123, BE2, BE14, BE31 or BE445 complex. The antibodies can be used to
CC antibody can be used to identify a patient having or at risk of
CC developing cancer
XX
SQ Sequence 515 AA;
Query Match 23.9%; Score 2585; DB 2; Length 515;
Best Local Similarity 99.0%; Pred. No. 5.4e-179;
Matches 510; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1630 NISWKKTIIVTRFLKLVDPDLAIIVQRKKGEERQAINROTALYTLKLCNFGAENPDPF 1689
Db |||||
1 NISWKKTIIVTRFLKLVDPDLAIIVQRKKGEERQAINROTALYTLKLCNFGAENPDPF 60
Qy 1690 VPLVXTAVKLIAPERKEKNVLSALLCTAEVSTLEALAIIPOLPSLMPSLTTMKNTSE 1749
Db |||||
61 VPLVXTAVKLIAPERKEKNVLSALLCTAEVSTLEALAIIPOLPSLMPSLTTMKNTSE 120
Qy 1750 LVSSEVYLLSALAALQKVETLPHFISPYLEGILSQVHLEKITSEMGSASQANIRLTSL 1809
Db |||||
121 LVSSEVYLLSALAALQKVETLPHFISPYLEGILSQVHLEKITSEMGSASQANIRLTSL 180
Qy 1810 KKTLATTLAPRVLLPAIKTYKQIEKNWNHMGPFMSILOEHIGXMKKEBELTSHQSOLTA 1869
Db |||||
181 KKTLATTLAPRVLLPAIKTYKQIEKNWNHMGPFMSILOEHIGXMKKEBELTSHQSOLTA 240
Qy 1870 FFLEALDFRAQHSNDELEVGTENCIIIDCLVAMVVKLSEVTRPLFFKLFWDKTEDAP 1929
Db |||||
241 FFLEALDFRAQHSNDELEVGTENCIIIDCLVAMVVKLSEVTRPLFFKLFWDKTEDAP 300
Qy 1930 KORLLTFYNLADCI AEKLGFLTFAGHLVKPPADTLXQVNI SKTDEAFFDSENDEPKCC 1989
Db |||||
301 KORLLTFYNLADCI AEKLGFLTFAGHLVKPPADTLXQVNI SKTDEAFFDSENDEPKCC 360
Qy 1990 LLLQFTLNCLYKIFLFDTOHFISKERAXALMPLVDQLENRLGGEKFOERVTKHLIPCI 2049
Db |||||
361 LLLQFTLNCLYKIFLFDTOHFISKERAXALMPLVDQLENRLGGEKFOERVTKHLIPCI 420
Qy 2050 AQFSVAMADDSLWKPLNYOILLKTRDSSPKVRPAALITVLALAEKLENYIVLLPESIPF 2109
Db |||||
421 VQFSVAMADDSLWKPLNYOILLKTRDSSPKVRPAALITVLALAEKLENYIVLLPESIPF 480
Qy 2110 LAELMEDECEEVHQCKTIQOLETVLGEPLQSYF 2144
Db |||||
481 LAELMEDECEEVHQCKTIQOLETVLGEPLQSYF 515

Search completed: November 7, 2005, 09:36:13
Job time : 248 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 7, 2005, 09:31:57 ; Search time 262 Seconds
(without alignments)
4190.450 Million cell updates/sec

Title: US-09-603-665-5
Perfect score: 10807
Sequence: 1 MTSLAQQLRLALPOSDASL.....CKTKIQQLTVLGEPLQSYF 2144

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Uniprot_03: *
1: uniprot_sprot: *
2: uniprot_trembl: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query Length	ID	Description
1	10799	99.9	2144	1 BP28_HUMAN	Q9h583 homo sapien
2	5807.5	53.7	2159	2 Q7SY48	Q7sy48 brachydanio
3	5562	51.5	1106	2 Q6P197	Q6p197 homo sapien
4	4729	43.8	958	1 BP28_MACFA	Q9gm44 macaca fasc
5	4494	41.6	897	2 Q8N7L7	Q8n7l7 homo sapien
6	3935.5	36.4	1336	2 Q7T152	Q7t152 brachydanio
7	3669.5	34.0	1278	2 Q7T153	Q7t153 brachydanio
8	3099	28.7	733	2 Q8BLJ4	Q8blj4 mus musculus
9	2094.5	19.4	2098	2 Q7PMD6	Q7pmd6 anopheles g
10	2051.5	19.0	2096	1 BP28_DROME	Q9vm75 drosophila
11	2051.5	19.0	2096	2 Q6AWS0	Q6aws0 drosophila
12	1926	17.8	408	2 Q8CCT5	Q8cct5 mus musculus
13	1779	16.5	349	2 Q96ES5	Q96es5 homo sapien
14	1682	15.6	349	2 Q8VCK1	Q8vck1 mus musculus
15	1462.5	13.5	1690	2 Q8T9E7	Q8t9e7 drosophila
16	1194.5	11.1	1830	1 BP28_ARATH	Q9c8z4 arabidopsis
17	1127	10.4	1857	2 Q6BXQ6	Q6bxq6 debaryomyce
18	1030.5	9.5	1788	2 Q7R2M8	Q7r2m8 neurospora
19	1006.5	9.3	1770	2 Q6FT93	Q6ft93 candida gla
20	1003.5	9.3	1774	2 Q6CUJ57	Q6cuj57 kluyveromyc
21	990.5	9.2	1649	1 YG44_SCHPO	Q60179 schizosacch
22	956	8.8	1769	1 YJK9_YEAST	P42945 saccharomyc
23	950.5	8.8	1774	2 Q6CAJ78	Q75418 ashbya goss
24	903	8.4	1635	2 Q6C457	Q6c457 yarrowia li
25	857	7.9	1650	1 BP28_CAEEL	Q23495 caenorhabdi
26	800	5.6	120	2 Q6P664	Q6p664 homo sapien
27	410.5	3.8	2381	2 Q7R4R8	Q7r4r8 giardia lam
28	365.5	3.4	2574	2 Q6BFG2	Q6bfg2 parametium
29	320.5	3.0	2736	2 Q6BMQ6	Q6bmq6 debaryomyce
30	317.5	2.9	2670	1 YAQ5_SCHPO	Q10105 schizosacch
31	309.5	2.9	2230	1 GOR4_HUMAN	Q13439 homo sapien

32 305 2.8 2671 2 Q754A3 Q754a3 ashbya goss
33 295 2.7 5058 2 Q86UQ4 Q86uq4 homo sapien
34 294 2.7 2792 2 Q814R2 Q814r2 plasmodium
35 293 2.7 3130 2 Q81DX6 Q81dx6 plasmodium
36 293 2.7 3254 2 Q7YWE9 Q7ywe9 plasmodium
37 293 2.7 3254 2 Q9BK45 Q9bk45 plasmodium
38 292 2.7 3130 2 Q9BK46 Q9bk46 plasmodium
39 291.5 2.7 3080 2 Q7YWF0 Q7ywf0 plasmodium
40 291.5 2.7 3081 2 Q7YWF1 Q7ywf1 plasmodium
41 291.5 2.7 3203 2 Q7YWE6 Q7ywe6 plasmodium
42 291.5 2.7 3203 2 Q7YWE7 Q7ywe7 plasmodium
43 290.5 2.7 2866 2 Q6FW99 Q6fw99 candida gla
44 290 2.7 3096 2 Q7YWF2 Q7ywf2 plasmodium
45 290 2.7 3256 2 Q7YWE8 Q7ywe8 plasmodium

ALIGNMENTS

RESULT 1
BP28_HUMAN
ID BP28_HUMAN STANDARD; PRT; 2144 AA.
AC Q9H583; Q9NW23;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 05-JUL-2004 (Rel. 44, Last annotation update)
DE Protein BAP28.
GN Name=BAP28;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND VARIANTS SER-1694; ALA-1854; ASP-1967 AND
RP GLY-2017.
RA Bougueret L., Chumakov I., Barry C., Cohen-Akenine A.;
RT "A novel BAP28 gene and protein.";
RL Patent number WO0100669, 04-JAN-2001.
RN [2]
RP SEQUENCE OF 1534-2144 FROM N.A.
RA Cobley V.;
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 1777-2144 FROM N.A.
RX PubMed=14702039; DOI=10.1038/ngl1285;
RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,
RA Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,
RA Sekine M., Oobayashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,
RA Yamamoto J.-I., Saito K., Kawai Y., Isono Y., Nakamura Y.,
RA Nagahari K., Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M.,
RA Shiratori A., Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H.,
RA Sugawara M., Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E.,
RA Omura Y., Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M.,
RA Yamazaki M., Ninomiya K., Ishibashi T., Yamashita H., Murakawa K.,
RA Fujimori K., Tanai H., Kimata M., Watanabe M., Hiraoa K., Chiba Y.,
RA Ishida S., Ono Y., Takiguchi S., Watanabe S., Yosida M., Hotuta T.,
RA Kusano J., Kanehori K., Takahashi-Fujii A., Hara H., Tanase T.-O.,
RA Nomura Y., Taguchi S., Komai F., Hara R., Takeuchi K., Arita M.,
RA Imose N., Musashino K., Yuuki H., Oshima A., Sasaki N., Aotsuka S.,
RA Yoshikawa Y., Matsunawa H., Ichihara T., Shiohata N., Sano S.,
RA Moriya S., Momiyama H., Satoh N., Takami S., Terashima Y., Suzuki O.,
RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,
RA Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,
RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,
RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,
RA Ono T., Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y., Okamoto S.,
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Senba T.,
RA Okitani R., Kakimoto T., Noguchi S., Itoh T., Shigeta K., Senba T.,
RA Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,
RA Togashi T., Oyama M., Hata H., Watanabe M., Komatsu T.,
RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,
RA Okumura K., Nagase T., Nomura N., Kikuchi Y., Masuho Y., Yamashita R.,
RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.;

Db 1621 DLNNKQQNISWKKYIVTRFLKLVDPDLAIVQKKKEGEQAINRQATYLTKLCKN 1680
Qy 1681 FGAENPDPPVPLVTKLAVKLIAPERKEKNVGLGALLCIAEVTSTLEBALAIPQLPMLPSL 1740
Db 1681 FGAENPDPPVPLVTKLAVKLIAPERKEKNVGLGALLCIAEVTSTLEBALAIPQLPMLPSL 1740
Qy 1741 LTTMKNTSELVSSEVILLSALAAKQKVETLPHFISPYLSEGLISQVHLEKITSSEMSAS 1800
Db 1741 LTTMKNTSELVSSEVILLSALAAKQKVETLPHFISPYLSEGLISQVHLEKITSSEMSAS 1800
Qy 1801 QANIRLTSKKTATLAPRVLPAIKTYKQLEKNKWHGPMSTLOSHICXMKKEEL 1860
Db 1801 QANIRLTSKKTATLAPRVLPAIKTYKQLEKNKWHGPMSTLOSHICXMKKEEL 1860
Qy 1861 TSHQSOLTAFFLEALDFAQHSNDLEEVGKTENCIIICLIVAMVVKLSEVTFPLPFKLF 1920
Db 1861 TSHQSOLTAFFLEALDFAQHSNDLEEVGKTENCIIICLIVAMVVKLSEVTFPLPFKLF 1920
Qy 1921 DWAKTEDAPKDRLLTFYNLADCTAEKLGFTLFAHGVKPPADTLXQVNIKTDEAFD 1980
Db 1921 DWAKTEDAPKDRLLTFYNLADCTAEKLGFTLFAHGVKPPADTLXQVNIKTDEAFD 1980
Qy 1981 SENDPEKCCLLQFILLNCLYKIFLFDTOHFISKERAXALMMPVLDQLENRLGGEKFOER 2040
Db 1981 SENDPEKCCLLQFILLNCLYKIFLFDTOHFISKERAXALMMPVLDQLENRLGGEKFOER 2040
Qy 2041 VTGHLPICTAQFSVAMADSLWKPLNYQILLKTRDSSPKVRFAALITVLALAEKLENYI 2100
Db 2041 VTGHLPICTAQFSVAMADSLWKPLNYQILLKTRDSSPKVRFAALITVLALAEKLENYI 2100
Qy 2101 VLLPESIPFLAEIEMEDECVEHQCKTIQOELTVLGEPLQSYF 2144
Db 2101 VLLPESIPFLAEIEMEDECVEHQCKTIQOELTVLGEPLQSYF 2144

RESULT 2

Q7SY48 PRELIMINARY; PRT: 2159 AA.
ID Q7SY48; AC Q7SY48;
DT 01-OCT-2003 (T-EMBLrel. 25, Created)
DT 01-OCT-2003 (T-EMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Hypothetical protein flj10359;
GN Names:flj10359;
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AB; TISSUE=Whole body;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McQuinn P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalek A., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences."

RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=AB; TISSUE=Whole body;
RA Director MGC Project;
RL Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC055128; AAH55128.1;
DR ZFIN; ZDB-GENE-030131-6378; flj10359.
DR InterPro; IPR008938; ARM.
DR InterPro; IPR000357; HEAT.
DR Pfam; PF02985; HEAT; 2.
KW Hypothetical protein.
SQ SEQUENCE 2159 AA; 242048 MW; F00DBBAID597E70B CRC64;
Query Match 53.7%; Score 5807.5; DB 2; Length 2159;
Best Local Similarity 53.8%; Pred. No. 3.5e-262;
Matches 1175; Conservative 391; Mismatches 554; Indels 63; Gaps 20;
Qy 1 MTSLAQOQLRALPOSDASILSRDEVASILLPDPKEAATIDRTAFAGCTGLEELGIDP 60
Db 1 MTSLAHQRLALPQNDSSLLGRKEVSVLPDPKDAASMDRSTFHALGCTGLEELMGIDA 60
Qy 61 SFEQPEAPLPSOLAKTLERSVOTKAVNKOLDENISLFLIHLSPYFLLKPAQKCLEWLIHR 120
Db 61 AFSEPOETLFSQSLTLERSVQSKVKNKDDKSISLFLTLSPYFLLKPAKLCIEWLHR 120
Qy 121 FHIHLNODSLACVLPYHETRI FVRVIOQLKINNKKHFWMLLPVKQSGVPLAKGTLT 180
Db 121 FHIHLNODSLACVLPYHETKVFVRVIOQLFKIEDTHKHWMLHGIOKPGVPLARGTLT 180
Qy 181 HCYKDLGFMDFCSSLVTKSVKFAEY--PGSSAQLRVLLAFYASTIVSALVAEVDSDNI 238
Db 181 HCYKDLGFMDFVCSMVTVNSKAYSELTRDGNCPQLRVIFSFYASTIVSALDAVEKITNSI 240
Qy 239 IAKLPYIOKGLKSSLPDYRAATYMIICQISVKVTMENTFVNSLASQIITLTKISLIK 298
Db 241 IAKLPFVQLGKSNLSLDYTAATYIVCMQVKNVMEQALVDSLSQLSRGLRTPOLR 300
Qy 299 DGLSLVLLQKQKPSLGKPPHLCNVPDLITILHGISETVDVSPLLRMYLPHLVVSI 358
Db 301 EGLSCIIILLQKQKGVIOKTYGYLCAVPTLVSTLQSIQSTVHDSPLSYLLPHLHVS 360
Qy 359 IHHVTGEETEGMDGQIYKRLHAILTKISKNNLDHLLASLLFFEEYISYSSOEBMSNKV 418
Db 361 MTQNDDEQNEGLSDS--TGLLSQSFQNLQSSNLENTAAKLLLEEVVVCN--ELPSDGI 416
Qy 419 SILNQFLPILLESKYPTLDVLEHLKEIADLKQELFHQFVSLSTSGCKYQFLAD 478
Db 417 SALNORIQTVRLEFSRYPCALDMALLENHVKVNSDNEKNLLHQFISLTSCKYQILPE 476
Qy 479 SDTSLMLSLHPLAPVRILAMNHLKKIMKTSKEGVDESFIKEAVLARLGDNDIDVVSAL 538
Db 477 SETSLMLSLHPLPSVRNMAVDYLKELNSENHNSFDEAFKDALLERIKDDSPVVSAL 536
Qy 539 SAFETPKHEFSSEVTISNLNLFQRAELSKNGEWELVIAADILIKELISLNDQLSNQ 598
Db 537 KALQHHMGLMDVEDTVSSLSILHR--IKPSADWCPVLKEAVRVLDPPRIEIGNPDLKAY 594
Qy 599 VVVCLLPFVIVNDDTESAEKIAIYLSKSGICSLHPLLRGHEEALENVISTKPKGLIG 658
Db 595 ISWELLFPFVIMTRAAPCEVELOMTSAITETTLISQHPLTQGWAKVLKAVLAKTSEDLIG 654
Qy 659 VANQKMIELADINILGPPSSMLKVEDLISVGESESFNLKOKVTFHVLISVLVSCSSL 718
Db 655 VANEMLTTLIKVLANMDHATKENTLENVCDILSRQSSVRDRAAFVVFSSALLQSLQM 714
Qy 719 KET-HFPFAIRVFSLLQKKIKKLESVITAVEITPSEWHIELMLDRGIPV-ELWAHYVEELN 776
Db 715 TESQHLHTAQSVYKLELLP-----LQAYTIQPEQVSQPADECLPVCVALGEFLQKIS 768
Qy 777 STQRVAVEDSVFLVFLSKKFIYALKAPKSPFGKDTWNPEQLKEDSRDYLHLLIGLFEMM 836
Db 769 CGLSAEQOGLLLSLRLFITTLKCPDSTFKGPMWNPKEKMTTTCYLRLLCLRFLDWW 828

```
Qy 837 1NGAD-----AVHFRVLMKFIKIVHLEDVQLPKFCVSLWTYSSUNPLNCSVKTVLOTQ 892
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 829 1SGASQGLPACFRSLMPQLQVHLNPMWLPFKFLSLSGWYNSNLGDLQDCRVSAILOQTQ 888
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 893 1LYVCCAMLSQKTOCKHQLASISPPVNTSLINLGSPKVEVRRRAIOCLQALSG-VASP 951
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 889 1ALYVGKAFSSQPVKTLNLLASDSPPVPSLLVVCVSGVCEVRRRAIAVQLCSLGVSSP 948
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 952 1FYLIIDHLISKAEEITSDAAYVIQDLATLIFEEL---OREKKLKHQKLSLSETLKNLLSCV 1007
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 949 1YHPIVEKLLKSSEIIADSSYLTAQLSKFEYEAHSVSKOKNKLAS-----VEQLLOCL 1001
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1008 1YS--CPSYIAKDLMKVLQVNGEMVLSQLLPMAEOLLEKIQEKPAPVAKDEAMVHLITLG 1065
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1002 1QSPFCPSYTKTLRALQDVHGEPVLSVLLPAVERLLEQCAPDSCTFLPDEALLQLLLS 1061
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1066 1KNEFSVSLNEDPKSLDIFIKAVHTTKELYAGMPTIOITALEKITKPEFAISDEKVQO 1125
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1062 1KFSMSAPLLVKDPRCLEVFIKALHTSARPYPTIPSFQITALEQITKPPFTAGDEKIQO 1121
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1126 1KLLRMLFDLLVNCNKHCAQTSSVSPKGISVNAEOVRIEPPDKAKPLGTVOQRRQKM 1185
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1122 1KILSLFDLLVGNKSPACQASINSVFKTIADVCELVANELIPADKQRTATVQOTRRSQM 1181
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1186 1OOKKSQDLR-SVOEVGSGYWQRTVLTLELLOHKKLRSPOILVPTLFNLLSRCLPPOE 1244
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1182 1--RKTQDTSAGVPEESVSWPRVTLLELLOHKKLKRQALVLPALFNLLSRCLPFAAE 1239
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1245 1QGMWYTKQILISCLLNIQKLSPDGKIPKQILDEKENVELIVOCIRLSMPQTHHA 1304
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1240 1QENIEYTKQILITCLNVCKLSPEGGPIKQVLEEDKFMELVQCVSRVSEMPQTHHA 1299
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1305 1LLLGVVAGIFDPKVLHNTMSIFTFMGANVRLDDTYSFQVINKTKVWIPALIOQSDSD 1364
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1300 1LLLGLAGLIPPEKVLHNTMPITFMGANIMRLDDTYSFQVINKTKVQAVIPALIKAEHG 1359
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1365 1SIEVSRNVBEIVKIIISVFVDALPHYPEHRRLPILVOLVDTLGAEPFLMILILLPEQVY 1424
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1360 1SSQSEGHMTVAQIIHVFVDALPHYPEHRRLPILSQLMSTLSPSRFLVWLMLLFPKHV 1419
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1425 1TKTVLAAVGEKDAILEADTEFWFSVCCPSVQHOQOSLMTLLOVLLKLPKEEETIPK- 1483
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1420 1TOTSAGATGAKEAVVERDQDFILVCCPEFVKEQTLKILQYLWLTLPQDREEAPEKK 1479
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1484 1---AVSFNKSSEOEAMLOVFNVEHTSKOLRHFKFLSVFMSQLSSNNFLKVV--RS 1537
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1480 1KPRGSAVKKDETVDL--IFSVEHSGKDLRHFKEFISIFMAQLLASDGFVGKVDACED 1537
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1538 1GSPFILLKGLBERLLETVLGYISAVAQSMERNADKLTVKFWRALSKAYDILLKVNALLPT 1597
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1538 1ITESTLQALQDILLVEVLYIQAVARVEDNADKPTAKFWRALLSKSYDTLQVNALPLM 1597
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1598 1ETIPVIRGLVGNPLPSVRKALDLNNKLOONISWKKTIIVTFKLVPLDILAIVORKKX 1657
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1598 1DTFITVMRGLVGNQLASVRKAMELNNKLQORTKMLKEQITALLELIGTLLSIVGRSHR 1657
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1658 1---EGEEEOAINRQATALYTKLLCKNFGAENPDPFVPVLTXTAKVIAPRKEEKNVLGS 1713
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1658 1QVTAQEEELAINRQATALYSLKLCRNFGSDHKEEFVPLVKNKAVELVA-DKDEEKNVMS 1716
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1714 1ALLCIAEVTSTLEALAIPOPLSLMPSLLTTMKNTSELVSSEVYLLSALAALQKVETPLH 1773
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1717 1ALLCVAEVTSTKALAIQPLHRLMPAVLDTLKERKOLLNNEIYLLSAVTLQRASETPLH 1776
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1774 1FISPYLSEGLISOVHLEKTISEMGSASQANIRLTSKKTATTLAPRVLPALKTKYKQI 1833
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1777 1FISPYLLDTLQVTRTLTLARRLTSCPLSVRLASLSTLTKLPPRVLLPTITKYCSM 1836
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1834 1EKNWKNMGPPWMSILOEHIGAMKKEELTSHSQSLTAPFFLEALDPAHQSENDLEVGKTE 1893
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1837 1VDAQONRLSPLMNLKHEHISHMDKQDLNNHQSLETSFFLSALDPFAHQCGQDLKKTAEIE 1896
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

```
Qy 1894 1NCIIDCLVAMVVKLSEVTFRPLFFKLPDWAKTEDAPKDRLLTFYNLADCIAEKLKGLFTL 1953
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1897 1GCVIDCLLWMKLUSEVTFRPLFFKLPDWSKIDGASKDRLLTFYRLADRIADKLKGLFVL 1956
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1954 1FAGHLVFPFADTLXQVNIKSTDEAFFDSNDPE-----KCCLLIQFIINCLYK 2001
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1957 1FAGLVKFPFSDLLHQLNISHTKAPFDSDESDSDSDDEADNDVTKSLLLYVLDCLHK 2016
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 2002 1FLPDTQHFISKERAXALMPLVDOLNRLGEGEKFOERVTKHLIPICIAQFSVAMADDSL 2061
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 2017 1FLYDTQHFISKERADALLCPLVDOLNMLGGEETYSRITTHLVPCIAQFAVAMRDSQ 2076
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 2062 1WKPLNYQILLKTRDSSPKVRFAALITVIALAELKENYIVLLPSIPFELABELMEDECEV 2121
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 2077 1WKVLYQILLKTRHSSPKVRFSALVMLELAGLRENYVMVLLPETIPFELABELMEDECEV 2136
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 2122 1EHQCKTITQOLETVLGEPLQSYF 2144
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 2137 1EHQVQVQIEMETILGEPLQSYF 2159
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
RESULT 3
Q6P197 PRELIMINARY; PRT: 1106 AA.
AC Q6P197
DT 05-JUL-2004 (TREMELrel. 27, Created)
DT 05-JUL-2004 (TREMELrel. 27, Last sequence update)
DT 05-JUL-2004 (TREMELrel. 27, Last annotation update)
DE FLJ10359 protein (Fragment).
GN Name=FLJ10359;
OS Homo sapiens (Human);
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RC SEQUENCE FROM N.A.
RP TISSUE=Lymph;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shremen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green B.D., Dickson M.C., Butterfield Y.S.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska J., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RC SEQUENCE FROM N.A.
RP TISSUE=Lymph;
RA Strausberg R.;
RL Submitted (JAN-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC065205; AAH65205.1; -.
DR InterPro; IPR008938; ARM.
DR InterPro; IPR000357; HEAT.
DR Pfam; PF02985; HEAT; 1.
FT NON_TER 1
SQ SEQUENCE 1106 AA; 125359 MW; 09F7CE94042302C4 CRC64;
Query Match 51.5%; Score 5562; DB 2; Length 1106;
Best Local Similarity 99.6%; Pred. No. 4e-251;
Matches 1102; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```


QY	1547	EBRLLETVLGYISAVAQSMERNADKLTVKFWRALLSKAYDLDKVNALLPTFTFPIRVG	1606
Db	360	EBRLLETVLGYINAVQAQSMERNADKLTVKFWRALLSKAYDLDKVNALLPTFTFPIRVG	419
QY	1607	LVGNPLPSVRRKALDNLNKLQNTSWKKTIVTRFLKLPDLDLAIIVORKKKEGEBEQAIN	1666
Db	420	LVGNPLPSVRRKALDNLNKLQNTSWKKTIVTRFLKLPDLDLAIIVORKKKEGEBEQAIN	479
QY	1667	ROTALYTLKLLCKNFGAENPDPPVPLXTAVKLIAPERKEKNVLSALLCIAEVTSTLE	1726
Db	480	ROTALYTLKLLCKNFGAENPDPPVPLXTAVKLIAPERKEKNVLSALLCIAEVTSTILO	539
QY	1727	ALAIPOPLSIMPSSLTMTKNTSELVSSEVYLLSALAALQKVETLPHFISPYLEGILSOV	1786
Db	540	ALAVPOLPSIMPSSLTMTKNTSELVSSEVYLLSALAALQKVETLPHFISPYLEGILSOV	599
QY	1787	IHLEKITSMSGSA-SQANIRLTSKKTATLAPRVLLPAIKTKYKQIEKNWNHMGPPM	1845
Db	600	IHLEKITSVGSASSQANIRLTSKKTATLAPRVLLPAIKTKYKQIEKNWNHMGPPM	659
QY	1846	SILQEHIGVWKKEELTSHOSQLTAFPLEALDFAHQHSENDLEEVGKTENCIIIDCLVAMVV	1905
Db	660	SILQEHIGVWKKEELTSHOSQLTAFPLEALDFAHQHSENDLEEVGKTENCIIIDCLVAMVV	719
QY	1906	KLSEVTFRPLFKLFDWAKTEDAPKDRLLTFYNLADCI AEKLGFTLPAFAGHLVKPFADT	1965
Db	720	KLSEVTFRPLFKLFDWAKTEDAPKDRLLTFYNLADCI AEKLGFTLPAFAGHLVKPFADT	779
QY	1966	LQVNIISKTDPAFFDSNDPEKCLLLOFLNCLYKIFLFDQHFISKERAALMPLVD	2025
Db	780	LQVNIISKTDPAFFDSNDPEKCLLLOFLNCLYKIFLFDQHFISKERAALMPLVD	839
QY	2026	QLENRLGGGEKFOERYTKHLIPICIAQFSVAMADDSLWKPLNYOILLKTRDSSPKVRFAAL	2085
Db	840	QLENRLGGGEKFOERYTKHLIPICIAQFSVAMADDSLWKPLNYOILLKTRDASPKVRFAAL	899
QY	2086	ITVLAALAEKLENYIVLLPESIPFLAEMLMEDECEVEHQCKTIOQLETVLGSPLOSIF	2144
Db	900	ITVLAALAEKLENYIVLLPESIPFLAEMLMEDECEVEHQCKTIOQLETVLGSPLOSIF	958
RESULT 5			
Q8NTL7			
ID	Q8NTL7	PRELIMINARY; PRT; 897 AA.	
AC	Q8NTL7		
DT	01-OCT-2002 (T-EMBLrel. 22, Created)		
DT	01-OCT-2002 (T-EMBLrel. 22, Last sequence update)		
DT	01-OCT-2003 (T-EMBLrel. 25, Last annotation update)		
DE	Hypothetical protein FLJ40893.		
OS	Homo sapiens (Human)		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TSUB=Uterus;		
RX	PubMed=14702039; DOI=10.1038/ng1285;		
RA	Ota T., Suzuki Y., Nishikawa T., Otauki T., Sugiyama T., Irie R.,		
RA	Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,		
RA	Sekine M., Obayashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,		
RA	Yamamoto J., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahara K.,		
RA	Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M., Shiratori A.,		
RA	Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,		
RA	Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E., Omura Y.,		
RA	Abe K., Kamihara K., Katsuma N., Sato K., Tanikawa M., Yamazaki M.,		
RA	Ninomiya K., Ishibashi T., Yamashita H., Murakawa K., Fujimori K.,		
RA	Tanai H., Kimata M., Watanabe M., Hiraoka S., Chiba Y., Ishida S.,		
RA	Ono Y., Takiguchi S., Watanabe S., Yosida M., Hotuta T., Kusano J.,		
RA	Kanehori K., Takahashi-Fujii A., Hara H., Tanase T., Nomura Y.,		
RA	Togiani S., Komai F., Hara R., Takeuchi K., Arita M., Imose N.,		
RA	Muashino K., Yuuki H., Oshima A., Sasaki N., Aotseuka S.,		
RA	Yoshikawa Y., Matsunawa H., Ichihara T., Shiohara N., Sano S.,		
RA	Moriya S., Momiyama H., Satoh N., Takami S., Terashima Y., Suzuki O.,		
RA	Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,		
RA	Highigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,		
RA	Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,		
RA	Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,		
RA	Ono T., Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y.,		
RA	Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,		
RA	Ohtani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,		
RA	Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Komatsu T.,		
RA	Togashi T., Oyama M., Hata H., Watanabe M., Takahashi Y., Nakagawa K.,		
RA	Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Yamashita R.,		
RA	Okumura K., Nagase T., Nomura N., Kikuchi H., Masuho Y., Sugano S.,		
RA	Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.,		
RT	"Complete sequencing and characterization of 21,243 full-length human		
RT	CDNAs."		
RL	Nat. Genet. 36:40-45 (2004).		
DR	EMBL; AK098212; BAC05261.1; --		
DR	InterPro; IPR008938; ARM.		
DR	InterPro; IPR000357; HEAT.		
PFam	PF02985; HEAT; 1.		
SQ	SEQUENCE 897 AA; 101574 MW; 5FF6A94FE8855895 CRC64;		
Query Match	41.6%; Score 4494; DB 2; Length 897;		
Best Local Similarity	99.2%; Pred. No. 2.1e-201;		
Matches 890; Conservative	2; Mismatches 5; Indels 0; Gaps 0;		
QY	1248	MEYTKQLILSCLNLT COKLSPDGGKIPKIDLBEEKFNVELIVQCIRLSEMPQTHHALLL	1307
Db	1	MEYTKQLILSCLNLT COKLSPDGGKIPKIDLBEEKFNVELIVQCIRLSEMPQTHHALLL	60
QY	1308	LGTVAGIPDKVHLNIMSI FTFMGANVMRLDDTYSFQVINKTVKMVIPALIOSDSGDSIE	1367
Db	61	LGTVAGIPDKVHLNIMSI FTFMGANVMRLDDTYSFQVINKTVKMVIPALIOSDSGDSIE	120
QY	1368	VSRNVEEIVVKIISVFDALPHVPEHRLPILVQLVDTLGAEKFLWILLILFEQVVTKT	1427
Db	121	VSRNVEEIVVKIISVFDALPHVPEHRLPILVQLVDTLGAEKFLWILLILFEQVVTKT	180
QY	1428	VLAAYGEKDALEADTEFWSCFVSQHOQISLMNITQVLLKLPESKEETIPKAVSF	1487
Db	181	VLAAYGEKDALEADTEFWSCFVSQHOQISLMNITQVLLKLPESKEETIPKAVSF	240
QY	1488	NKSEOEEMLOVFNVEHTS KQLRHFPLSVFMSQLSSNNPILKVVESGGPEILKGLGE	1547
Db	241	NKSEOEEMLOVFNVEHTS KQLRHFPLSVFMSQLSSNNPILKVVESGGPEILKGLGE	300
QY	1548	ERLLETVLGYISAVAQSMERNADKLTVKFWRALLSKAYDLDKVNALLPTFTFPIRVGL	1607
Db	301	ERLLETVLGYISAVAQSMERNADKLTVKFWRALLSKAYDLDKVNALLPTFTFPIRVGL	360
QY	1608	VGNPLPSVRRKALDNLNKLQNTSWKKTIVTRFLKLPDLDLAIIVORKKKEGEBEQAINR	1667
Db	361	VGNPLPSVRRKALDNLNKLQNTSWKKTIVTRFLKLPDLDLAIIVORKKKEGEBEQAINR	420
QY	1668	Q7ALYTLKLLCKNFGAENPDPPVPLXTAVKLIAPERKEKNVLSALLCIAEVTSTLEA	1727
Db	421	Q7ALYTLKLLCKNFGAENPDPPVPLXTAVKLIAPERKEKNVLSALLCIAEVTSTLEA	480
QY	1728	LAIPQLPSLMPSSLTMTKNTSELVSSEVYLLSALAALQKVETLPHFISPYLEGILSOVI	1787
Db	481	LAIPQLPSLMPSSLTMTKNTSELVSSEVYLLSALAALQKVETLPHFISPYLEGILSOVI	540
QY	1788	HLEKITSMSGSAQANIRLTSKKTATLAPRVLLPAIKTKYKQIEKNWNHMGPPMSI	1847
Db	541	HLEKITSMSGSAQANIRLTSKKTATLAPRVLLPAIKTKYKQIEKNWNHMGPPMSI	600
QY	1848	LQEHIGVWKKEELTSHOSQLTAFPLEALDFAHQHSENDLEEVGKTENCIIIDCLVAMVVKL	1907
Db	601	LQEHIGVWKKEELTSHOSQLTAFPLEALDFAHQHSENDLEEVGKTENCIIIDCLVAMVVKL	660
QY	1908	SEVTFRPLFKLFDWAKTEDAPKDRLLTFYNLADCI AEKLGFTLPAFAGHLVKPFADTIX	1967
Db	661	SEVTFRPLFKLFDWAKTEDAPKDRLLTFYNLADCI AEKLGFTLPAFAGHLVKPFADTIX	720

```

QY 1968 QVNISKTDAPFDSNDPEKCCLLQFILNCLYKIFLDTQHFISKERAXALAMPLVDQL 2027
DB 721 QVNISKTDAPFDSNDPEKCCLLQFILNCLYKIFLDTQHFISKERAGALAMPLVDQL 780
QY 2028 ENRLGGEKEQERVTHLIPCIAPQSVAMADDSLWKPLNYQILLKTRDSSPKVRFAALIT 2087
DB 781 VNRLLGGEKEQERVTHLIPCIAPQSVAMADDSLWKPLNYQILLKTRDSSPKVRFAALIT 840
QY 2088 VLALAEKIKENYIVLLPESIPFLAELMEDECEVEHQCCKTIQOELTVLGEPLQSYF 2144
DB 841 VLALAEKIKENYIVLLPESIPFLAELMEDECEVEHQCCKTIQOELTVLGEPLQSYF 897

RESULT 6
Q7T152 PRELIMINARY; PRT; 1336 AA.
AC Q7T152;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DE SI:zC146F4.2.1 (Novel protein similar to human BAP28) (Fragment).
GN Name=SI:zC146F4.2;
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN 1
RP SEQUENCE FROM N.A.
RA Garner P.;
RL Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL732629; CAE17603.1; -
DR InterPro; IPR008938; ARM.
FT NON TER 1
SQ SEQUENCE 1336 AA; 150326 MW; AA949557F21ACBCE CRC64;

Query Match 36.4%; Score 3935.5; DB 2; Length 1336;
Best Local Similarity 58.6%; Pred. No. 4.3e-175;
Matches 794; Conservative 220; Mismatches 297; Indels 45; Gaps 13;

QY 812 WNNPEOLKEDSRDYLHLLIGLPEMMLNGAD---AVHFRVLMKLFKIVHLEDVQFLPKFC 867
DB 3 WNNPEOLKEDSRDYLHLLIGLPEMMLNGAD---AVHFRVLMKLFKIVHLEDVQFLPKFC 62
QY 868 SVLWYTGSSILNPLNGSVKTVLTQALYVGCAMLSQKTCCKHQLASISPPVTSLLNL 927
DB 63 SVLWYTGSSILNPLNGSVKTVLTQALYVGCAMLSQKTCCKHQLASISPPVTSLLNL 122
QY 928 GSPVKEVRRAAIOCLQALSG-VASPPYLIIDHLISKAEITSDAAYVIOQLATLPEEL-- 984
DB 123 GSPVKEVRRAAIOCLQALSG-VASPPYLIIDHLISKAEITSDAAYVIOQLATLPEEL-- 182
QY 985 --QREKKLKHQKLSLTKNLLSCVYS--CPSYIAKLMKVLQGVNGEVMVLSQLPMAEQ 1040
DB 183 RDXDNKKKLAS-----VQLQCLQSPFCPSYTSKTLRLAQDVHGEFVSLVPAVER 235
QY 1041 LLEKIQKEPTAVLKDEAMVHLHTLGYKNFVSLSLNEDEPKSLDIFTKAVHTTKELVAGMP 1100
DB 236 LLEKIQKEPTAVLKDEAMVHLHTLGYKNFVSLSLNEDEPKSLDIFTKAVHTTKELVAGMP 295
QY 1101 TQITALEKIKTPFFAAISDEKVVQKLLRMLFDLLVNCNKHCAQTQVSSVFKGISVNAEQ 1160
DB 296 SFQITALEKIKTPFFAAISDEKVVQKLLRMLFDLLVNCNKHCAQTQVSSVFKGISVNAEQ 355
QY 1161 VRIELEPPKAPLGVQKRRQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQK 1219
DB 356 VRIELEPPKAPLGVQKRRQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQK 413
QY 1220 LRSPQILVPTLNNLSRCLLEPLPQEQNMMEYTKQLILSCLLNI COKLSPDGKIPKIDLD 1279
DB 414 LRSPQILVPTLNNLSRCLLEPLPQEQNMMEYTKQLILSCLLNI COKLSPDGKIPKIDLD 473

```

```

QY 1280 EERFNVELIVQICIRLSEMPQTHHALLLLCTVAGIFPDKVLVHNMISFTFMGANVMDLD 1339
DB 474 EERFNVELIVQICIRLSEMPQTHHALLLLCTVAGIFPDKVLVHNMISFTFMGANVMDLD 533
QY 1340 TYSFQVINTVKVVIIPALIOSDSDSIEVSERNVEEIVKIIISVFDVDPALPHVPEHRRPIL 1399
DB 534 TYSFQVINTVKVVIIPALIOSDSDSIEVSERNVEEIVKIIISVFDVDPALPHVPEHRRPIL 593
QY 1400 VQLVDLTGAEKFKFLLIILFEQVTVTKTLLAAAYGEKDAILEADTEFWFSVCCFVSQHQ 1459
DB 594 VQLVDLTGAEKFKFLLIILFEQVTVTKTLLAAAYGEKDAILEADTEFWFSVCCFVSQHQ 653
QY 1460 IQSLMNILOYLLKLPEKESETIPK-----AVSFNKGESQBEMLQVFNVTHTSKQLRHF 1514
DB 654 IQSLMNILOYLLKLPEKESETIPK-----AVSFNKGESQBEMLQVFNVTHTSKQLRHF 711
QY 1515 FLSVSFMSQLLSSNNFLKKVV--ESGPEILKGLERLLETVLGYISAVAQSMERNADKL 1572
DB 712 FLSVSFMSQLLSSNNFLKKVV--ESGPEILKGLERLLETVLGYISAVAQSMERNADKL 771
QY 1573 TVKFWALLSKAYDILLDKVNALLPETFTPIVIRGLVGNPLPSVRRKALDLNNKLOQNIS 1632
DB 772 TVKFWALLSKAYDILLDKVNALLPETFTPIVIRGLVGNPLPSVRRKALDLNNKLOQNIS 831
QY 1633 WKTIIVTRFLKLVDPDLAIIVQRKK---EGEBEQAINRQTALYTLKLCKNFGAENPDP 1688
DB 832 WKTIIVTRFLKLVDPDLAIIVQRKK---EGEBEQAINRQTALYTLKLCKNFGAENPDP 891
QY 1689 FVPVLTAVKLTAPERKEKKNVLSGALLCIAEVTSTLEALAIIPQLSLMPSLITTKWNTS 1748
DB 892 FVPVLTAVKLTAPERKEKKNVLSGALLCIAEVTSTLEALAIIPQLSLMPSLITTKWNTS 950
QY 1749 ELVSVSEVYLLSALAAQKVVETLPHFISPVLEIGILSQVHLHEKITSEMGSASQANRLTS 1808
DB 951 ELVSVSEVYLLSALAAQKVVETLPHFISPVLEIGILSQVHLHEKITSEMGSASQANRLTS 1010
QY 1809 LKKTTLATLAPRVLLPAIKTYKQIEKNWNHMGPFMSILOEIGHMKKEBELTSHOSQLT 1868
DB 1011 LKKTTLATLAPRVLLPAIKTYKQIEKNWNHMGPFMSILOEIGHMKKEBELTSHOSQLT 1070
QY 1869 AFPLEALDFAQHSNDLEEVGKTENCIIDCLVAMVVKLSEVTFRPLFPKLFQWAKTEDA 1928
DB 1071 AFPLEALDFAQHSNDLEEVGKTENCIIDCLVAMVVKLSEVTFRPLFPKLFQWAKTEDA 1130
QY 1929 PKDRLLTFYNLADCAEKLGLFTLPAHLVVKPFADTLQVNIKSKTDEAFDSDENPEKC 1988
DB 1131 PKDRLLTFYNLADCAEKLGLFTLPAHLVVKPFADTLQVNIKSKTDEAFDSDENPEKC 1180
QY 1989 CLLLOFILNCLYKIFLDTQHFISKERAXALAMPLVDQLENRLGGEKEQERVTHLIP 2048
DB 1181 CLLLOFILNCLYKIFLDTQHFISKERAXALAMPLVDQLENRLGGEKEQERVTHLIP 1240
QY 2049 IAQFSVAMADDSLWKPLNYQILLKTRDSSPKVRFAALITVLALAEKIKENYIVLLPESIP 2108
DB 1241 IAQFSVAMADDSLWKPLNYQILLKTRDSSPKVRFAALITVLALAEKIKENYIVLLPESIP 1300
QY 2109 FLAELMEDECEVEHQCCKTIQOELTVLGEPLQSYF 2144
DB 1301 FLAELMEDECEVEHQCCKTIQOELTVLGEPLQSYF 1336

RESULT 7
Q7T153 PRELIMINARY; PRT; 1278 AA.
AC Q7T153;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DE SI:zC146F4.2.2 (Novel protein similar to human BAP28) (Fragment).
GN Name=SI:zC146F4.2;
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;

```

OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]_SEQUENCE FROM N.A.
RP SEQUENCE FROM N.A.
RA Garner P.;
RL Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL732629; CAB17602.1; -;
FT InterPro; IPR008938; ARM.
FT NON TER 1
SQ SEQUENCE 1278 AA; B6C9FC81B77EE1A9 CRC64;

Query Match 34.0%; Score 3669.5; DB 2; Length 1278;
Best Local Similarity 57.6%; Pred. No. 1e-162;
Matches 744; Conservative 211; Mismatches 292; Indels 45; Gaps 13;

QY 812 WNPPEOLKEDSDYLLHLGLFEMMLNGAD----AVHFRVLMKLFIKVHLEDFVQLFKPC 867
DB 3 WNPPEKMETTCCYLRLLCCLFDVVISGASQGPLAPCFRSLMQPLLQVHLNPMWLFKFL 62
QY 868 SVLWTVGSSLSNPLNSVTVTLQALYVGCAMLSQKTCQKHQLASISSPVVTSLLINL 927
DB 63 SLLGWYNSNLGDLDCRVSAILQALYVYKAPFLSSQPVKTLNLLASDSSPVVPSLLVCV 122
QY 928 GSPVEVRRAATOCALSG-VASPEYLIIDHLISKAERITSDAAVVIDLATLPEEL-- 984
DB 123 CSGVCEVRRAATAVLQCLSLGVSSPVHPLVEKLKSGSEIIADSSVLTQALSKFYEEAVS 182
QY 985 --QREKKLSHOKLSETLKNLLSCVYS--CPSYIAKDLMKVLQGVNGEMVLSQLPMAEQ 1040
DB 183 RKDKNKKLAS-----VQLQCLQSPFCPSYTSKTLRALQDVHGEVPLSVLLPAVER 235
QY 1041 LLEKIQKEPTAVLKDAMVHLHLTLGKNBFSVSLNEDPKSLDIFIKAVHTTKELYAGMP 1100
DB 236 LLEQCAPDSCTFLPDEALLQLLSKFSMSAPLLVKDPRCLEVFTRALHTSARPYPTIP 295
QY 1101 TQITALEKTKPFFAAISDEKVOOKLMLPDLVLCNKSCHCAQTVSSVFKGISYNASQ 1160
DB 296 SFQITALEQTKPFFTAIGDEKIQOKLILFDLLVGNKSPACAAQINSVFKTIAVDCEL 355
QY 1161 VRTELEPPDKAKPLGTVQQRKQMOQKKSQDLLE-SVQEVGGSGYWRVTLILELLOHKKK 1219
DB 356 VANELIPADKQRTATVQOTRSEK--RKQDTSGAVPEESVVSFWRVTLILELLOHKKK 413
QY 1220 LRSPOILVPTFLNLSRCLPPLPOEQGNMYTKOLISCLLNTCOKLSPDGKIPKXILD 1279
DB 414 LKRAQVLPALFNLLSRCLPEAAAEQENIEYTKQLILICLLNVCKLSPGEGGPIKDVLE 473
QY 1280 EBFNVVELIVOCIRLSEMPOTHHALLLIGTVAGIPDKVLHNMISTFTFMGANVRLDD 1339
DB 474 EDKFNELVVQCRVSEMPQTHHALLLIGLALAGIPPEKVLHNMIPITFMGANVRLDD 533
QY 1340 TYSFQVINKTVKMWIPALIQSDGSDSTEVSRRNVEEIVVKIISVFVDALPHVPEHRLPIL 1399
DB 534 TYSFQVINKTVQAVIPALIKAHGSGSQSGHMETVVAQIIHVFVDALPHVPEHRLPIL 593
QY 1400 VQLVDTLGAKFLLWILLILPEQVTKTVLAAAYGEKDAILEADTEFWFVSCFESVQHQ 1459
DB 594 SQLMSTLGPFRFLVWMLLFLFKQHVQTQSAGATGAKEAEVVERDQDFWILLVCEFEVKEQ 653
QY 1460 IQSLMNTLOYLLKLPKEKETIPK----AVSFNKSQSEQLQVFNVEHTSKQLRHFK 1514
DB 654 LTSILKTLQVMTLPQDREAEPEKKPRGSVAVKQDFTVSDL--IFSVEHSGKOLRHFK 711
QY 1515 FLVSFMSQLSSNFFLKVV--ESGGPEILKGLERLLETVLGYISAVAQSMERNADKL 1572
DB 712 FISISFNAQLASDGFVGKADCEDITEITLQALQDQLLVEVLYRIQAVARCVEDNADKP 771
QY 1573 TVKFWRALSKAYDILLDKVNALLPTETFIPIVIRGLVGNPLPSVRRKALDILNKKLQNIS 1632
DB 772 TAKFWRALSKSYDTLQVKNALLPMDTFITVMRGLMGNQLASVRRKAMELNNKLQORTK 831
QY 1633 WKKTIVTRFLKLVPLDLALVORKKK---EGEEEOALNROTALYTLKLLCKNFCAENPDP 1688

832 WLKEQITALLELIGTLLSIVGRSHQVTAQEEBELAINROTALYSLKLLCRNFGSDHKEE 891
1689 FVPVLXTAVGLIAPERKEEKVNLGSALLCIAEVTSTLEALAIPOPLPSLMPSLTTMKNTS 1748
892 FVPVLNKAVELVA-DKDEKKNVMSALLCVAEVTSTLKALAIPQLHRLMPAVLDTLKERK 950
1749 ELVSESVYLLSALALQKVVETLPHFIPISYLEGILSQVHLEKITSFEMGASQANIRLTS 1808
951 DULNNEIYLLSAVTLAQASSETLPHFISPYLLDTLQVTRTLTLAARLTSCPLSVRLAS 1010
1809 LKKTTLATLAPRVLLPAIKTKYQIEKNVKNHMGFMFSILOSHIOMXMKKEELTSQSQLT 1868
1011 LSSTLATKLPREVLIPTITKCYCSMVDAQONRLSPLMNLKEHISHMDKQDLNNHQSELT 1070
1869 AFFLEALPRAQSENDEEVEGKTENCILDCIVAMVVKLSEVTFPPLPKLFDWAKTADA 1928
1071 SFPLSALDFRAHQCOGDLKKTAEIEGCVIDCLLVMMKMLSEVTFPPLPKLFDWAKTADA 1130
1929 PKDRLLTFVNLADCTAEKLGFLTFAGHLVKFPADTLXQVNNISKTDFAFFDSENDPEKC 1988
1131 SKDRLLTFYRLADRIADKLKGLFVLFAQOLVKFPFSDLLHQLNTSHTGKS-----S 1180
1989 CILLQFIILNCLYKIFLFDTHFISKERAXALMPLVDLENLIGREKQFQERTKHLIPC 2048
1181 SLLQVLDLCLHKIFLYDTHQFELSKERADALLCPLVDQLENLIGRETYKSRITTHLVCPC 1240
2049 IAQFSVAMADDSLWKPLNLYQIILLKTRDSSPKV 2080
1241 IAQFAVAMRDDSQWVLYNLYQIILLKTRHSSPKV 1272

RESULT 8
Q8BLJ4 PRELIMINARY; PRT; 733 AA.
AC Q8BLJ4;
DT 01-WAR-2003 (TrEMBLrel. 23, Created)
DT 01-WAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Mus musculus 9.5 days embryo parthenogenote cDNA, RIKEN full-length enriched library, clone: B130016L12 product: similar to PROTEIN BAP28 (Fragment).
DE Name: B130016L12Rik;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Parthenogenote;
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.,
RL Meth. Enzymol. 303:19-44 (1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Parthenogenote;
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA RIKEN FANTOM Consortium;
RL "Functional annotation of a full-length mouse cDNA collection."
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Parthenogenote;
RA The FANTOM Consortium,
RL "Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs."
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Parthenogenote;
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,

RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
 RT "Normalization and subtraction of cap-trapper-selected cDNAs to
 RT prepare full-length cDNA libraries for rapid discovery of new genes";
 RN Genome Res. 10:1617-1630(2000).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Parthenogenote;
 RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
 RA Shibata K., Itoh M., Aizawa K., Nagaoaka S., Sasaki N., Carninci P.,
 RA Konno H., Akiyama J., Nishi K., Kitsuana T., Tashiro H., Itoh M.,
 RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
 RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
 RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
 RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsushima S., Kawai J.,
 RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
 RT "RIKEN integrated sequence analysis (RISA) system-384-format
 RT sequencing pipeline with 384 multicapillary sequencer";
 RN Genome Res. 10:1757-1771(2000).
 RN [6]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Parthenogenote;
 RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
 RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
 RA Hayaehida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,
 RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
 RA Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
 RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
 RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,
 RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
 RA Sasaki D., Shibata K., Shingawa A., Shiraki T., Sogabe Y., Tagami M.,
 RA Tagawa A., Takahashi T., Takaku-Akashira S., Takeda Y., Tanaka T.,
 RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AK0494969; BAC32161.1; -
 DR MGD; MGI:2442524; B13001512Pik.
 DR InterPro; IPR008938; ARM.
 DR Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 DR Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Anopheles.
 DR Pfam; PF02985; HEAT; 2.
 DR NON_TER 733 733
 FT SEQUENCE 733 AA; 81938 MW; 513394872B6250EC CRC64;

Query Match 28.7%; Score 3099; DB 2; Length 733;
 Best Local Similarity 81.3%; Pred. No. 2.1e-136;
 Matches 596; Conservative 73; Mismatches 64; Indels 0; Gaps 0;
 QY 1 MTSLAQQLORLALPQSDASLLSRDEVASLFLDPKKAATIDRDTAFAGTCTGLEELGIDP 60
 Db 1 MTSLAQQLORLALPQSDASLLSRDEVASLFLDPKKAATIDRDTAFAGTCTGLEELGIDP 60
 QY 61 SPEQEPAPLPSQAKTLERSVQTKAVNKQDENISLFLHLSPYFLKPKAQKLEWLIHR 120
 Db 61 AFEQEPAPLPSQAKTLERSVQTKAVNKQDENISLFLHLSPYFLKPKAQKLEWLIHR 120
 QY 121 FHILHYNQSLACVLPYHETRIFFVRIQVLLKNNKSHRWFWLLPVKQSGVPLAKGTLT 180
 Db 121 FHILHYNQSLACVLPYHETRIFFVRIQVLLKNNKSHRWFWLLPVKQSGVPLAKGTLT 180
 QY 181 HCYKDLGFMDFICSLVTKSVKVPFAEYPGSSAQLRVLAFYASTIVSALVAEDVSNITA 240
 Db 181 HCYKDLGFMDFICSLVTKSVKVPFAEYPGSSAQLRVLAFYASTIVSALVAEDVSNITA 240
 QY 241 KLFPYIQKGLKSLPYDAATYMIICQISVKVTMENTFVNSLASQIIKTLTKIPSLIKDG 300
 Db 241 KLFPYIQKGLKSLPYDAATYMIICQISVKVTMENTFVNSLASQIIKTLTKIPSVQNDG 300
 QY 301 LSLVLLQRPESLGKPPHPLCNVPLITILHGISITYDVSPLLRMLPHLVSIITH 360
 Db 301 LGCLLIQRPENLGERPFLHLCVPLDILGHGISYDVSPLLRMLPHLVASVQ 360
 QY 361 HVTGETQMDGQIYKRLHLEILTKISLAKNNLDHLLALLFEEYISYSSQEEEDMSKYL 420
 Db 361 HIAGEBEAGIDGQIYKRLHLEILTKISLAKNNLDHLLALLFEEYISYSSQEEEDMSKYL 420

QY 421 LNEQFLPLRLRLSKYPRTLDVVLLEHLKEIADLKKQELPHQFVSLSTSGKQYQFLADSD 480
 Db 421 LNEQFLPLRLRLSKYPRTLDVVLLEHLKEIADLKKQELPHQFVSLSTSGKQYQFLADSD 480
 QY 481 TSLMLSLNHPPLAPVRIILAMNHLKIMKTSKEGVDESIFKEAVLARLGDNDIDVLSAISA 540
 Db 481 TSLMLSLNHPPLAPVRIILAMNHLKIMKTSKEGVDESIFKEAVLARLGDNDIDVLSAISA 540
 QY 541 FEIFKEHFSSEVTISNLLNLFQRAEILSKNGEWEVLKIAADILITKEILSENDOLSNOVV 600
 Db 541 FEIFQHQHFGVEETISNLLNLFQRAEILSKNGEWEVLKIAADILITKEILSENDOLSNOVV 600
 QY 601 VCLLPFWVINDDTESAEMKIAIYLSKSGICSLHPLLRGWEAELENVIKTKPKGLIGVA 660
 Db 601 VQLLPFWVITNSDIESPDMKIAIHLKSGICSLHPLLRGWEAELENVIKTKPKGLIGVA 660
 QY 661 NQKMIELLADNINLGPSSMLKQVEDLISVGEESFNLKQKVTTFHVLSVLVSCCSLKE 720
 Db 661 NQKMWQLLGSNLGSRSTVLKVEDLVACAGEKESYSKQKVAHFVTVSVLISCCSFQE 720
 QY 721 THPPFAIRVFSLL 733
 Db 721 TCPPFAIRVFSLL 733
 RESULT 9
 Q7P6D6 PRELIMINARY; PRT; 2098 AA.
 ID Q7P6D6 PRELIMINARY; PRT; 2098 AA.
 AC Q7P6D6; 01-MAR-2004 (TREMELrel. 26, Created)
 DT 01-MAR-2004 (TREMELrel. 26, Last sequence update)
 DT 01-MAR-2004 (TREMELrel. 26, Last annotation update)
 DE ENSANGP00000017181 (Pragmat).
 GS Name=ENSANGP00000014692;
 GN Anopheles gambiae str. PEST.
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Anopheles.
 OX NCBI_TaxID=180454;
 RN [1]_-
 RP SEQUENCE FROM N.A.
 RC STRAIN=PEST;
 RA Anopheles Genome Sequencing Consortium;
 RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
 CC -!- CAUTION: The sequence shown here is derived from an
 CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
 CC preliminary data.
 DR EMBL; AAAB01008984; BAA14843.1; -
 DR InterPro; IPR001917; Aminotrans_II.
 DR InterPro; IPR008938; ARM.
 DR InterPro; IPR000357; HEAT.
 DR Pfam; PF02985; HEAT; 3.
 DR PROSITE; PS00599; AA_TRANSFER_CLASS_2; UNKNOWN_1.
 FT NON_TER 1 1
 SQ SEQUENCE 2098 AA; 235889 MW; F4ED8C24CEAD0070 CRC64;
 Query Match 19.4%; Score 2094.5; DB 2; Length 2098;
 Best Local Similarity 28.9%; Pred. No. 6.8e-89;
 Matches 667; Conservative 401; Mismatches 859; Indels 377; Gaps 69;
 QY 2 TSLAQLQRLALPQSDASLLSRDEVASLFLDPKKAATIDRDTAFAGTCTGLEELGIDPS 61
 Db 11 TGLAQLKRLAAPQTSAPVDAR-RTASILFDKAEAAKDRVYIDIGVSGLEELTQMHGA 69
 QY 62 FQEPAPLPSQAKTLERSVQTKAVNKQDENISLFLHLSPYFLKPKAQKLEWLIHRP 121
 Db 70 FAQFEDTLFDKAMDOLQSRVENREYQDGNIRRFVHLSPYFLMQLPAKLEWLIHRP 129
 QY 122 HHLHYNQSLACVLPYHETRIFFVRIQVLLKNNKSHRWFWLLPVKQSGVPLAKGTLT 181
 Db 130 EHLHYNRELFWLILFPFHETRIFFVRCVQTMQLTEEVNDRILAAVKSQVPAKRTI 189
 QY 162 CYKDLGFMDFICSLVTKSVKVPFAEYPGSSAQLRVLAFYASTIVSALVAEDVSNITAK 241

Db 190 CVSHGFLQXQSFITGAVE---ELAGRANALQATFAPYCTATGMLHSAETVSENHVTA 246
Qy 242 LFPYIQKGLSLPYRAATYMIICQISVYKVTMENTFVNSLASQIIKTITKIPSLIKDGL 301
Db 247 VLHTVCKGLASRAIDPAAGSFVIVQVLVWKASLAQLTVDYLARRVI--AVQLPALTAET 304
Qy 302 SCLIVLLQRO---KPSISG---KKPPPHLCNVPDDLITILHGISYDVVSPLRY 349
Db 305 MLLVLIFQOHERLATLSRELMEIRRCCKWLAPTLCVK-----ADGVDVLVLYRK 355
Qy 350 MFLPHVSVSIHHVTGEETGMDGOIYKRHLAILTKISIKNNLDHLLASLLFEEY---- 404
Db 356 LLEKCLNEICC-----TKGA-LKLYSGFCEOLMUEIQUTEVEAEVIVQCVLDVSFHKDV 408
Qy 405 -----ISYSSQEMD-SNKVSLNQEQLPLIRLLESKYPRTLDVVLEHLK- 449
Db 409 PEKKAANSDDTIELDSEEDDFVSRDOHVQWYSEVYKSPEROYPGAFAKVVVKRMKG 468
Qy 450 -BIADLKQELPHQFVSLSTSGGKYQFADSDTSLMLSNHPLAPVRILAMNHL---KKI 505
Db 469 QEOYSPKKNALRNVLGFLLOAS-----VDENETNVPESLFHYDADRRAHVOYLVENLRA 524
Qy 506 MKTSEKGVDESGFIEKAVLARLGDNDIVV--LSAISAFIEIKEHFSSEVTISNLLNFQ 562
Db 525 MKLKSTGQVD--LLRDSVRERLADSCVEIBELKIDPOELIVVGTDEL-IGRLSQLVL 581
Qy 563 RAELSKNGEWEVLKIAADILIKEEILSENDSQSNQVVCILPFPV--INN- 612
Db 582 KCAANQT-RWAKTPRVIELLDRVYVYGRGN--VNOIITIALYPLFPIGNDPVGORNKA 638
Qy 613 -----DTESAMKIAIYLSKGICSLHPLLRGWEALENV--IKSTKP 653
Db 639 LLSTSPAKOYLVARFYNGEENCLEMLSTLL-BAGQCT-NPV-----QVCFNILLSSALP 691
Qy 654 GK-LIGVANQKM-----TELLADNINLGDPPSMLKVEDLISVGE 693
Db 692 TPCVTGQADRVLDYAVRQLQNHFRFYAPTSDIQCLQDN-----QLPODLTLVPIR 741
Qy 694 ESNLAKQKTFHVLVS-LVSCSSLSKETHFPFAIRVFSLOKKIKKLESVITAVEIPSE 752
Db 742 ---YILERVQOPEIGANFRQCVSLK-----LRLAIFAVILLEOYCTIDPOHGA- 788
Qy 753 WHIELMDRGIPVELMAHYVEELNSTORVAVESVFLVPSLKKFIYALKAPKSPFKGDW 812
Db 789 -----RSIFGELLKSYLSRYPQ---LVDRIEFL-----SHF 817
Qy 813 WNEQKEDSRDYLHLLIGLFEMWLNAGADAVHFRVLMKLFIKVHLEDVFLFKFCVSLWT 872
Db 818 YTAHLIEPGSKD-----DADNFLINAKQIR-----CORL-- 847
Qy 873 YGSSLSNPLNCSVKTVLQALYVGCAMLSSQTKQCKHOLASTSSPVVTSLLINLGPVK 932
Db 848 -----LMTLEQVSQES---GC-----TVTESALINVLISLTSGSA 881
Qy 933 EVRAATQICLQAL-----SGVASPFYLIIDHISKAETSAAAYIQDLA-TLFEELQ 985
Db 882 VRECTMRTLIQIVGLKRNGLSKAFNGFVKRMVYKREELTMDG-----EQLALWFAIFS 937
Qy 986 REKKLSHQKLSLTKNLLSCVY--SCPSYIAKDLMKVLOGVNGEMVLSQLLPWASQOLLE 1043
Db 938 LEEBEGDGLPQPMESFVARIVDPSTPTIYVCGIVDLLKLLDDPPHLLSCASKGVSIIE 997
Qy 1044 KIQKEPTAVLKD--EAMVLHILTKGNEFSVSLINEDPKSLDIPKIAVHTTKELYAGMP- 1100
Db 998 RVNAAASQPVDAVESRVVQLLITRFLNLDIVAHLPGSALCORLVTLAIRCCKPLC--LPS 1055
Qy 1101 -----TIQITALEKITPPFAAISDEKVOQKLRMLFDLIVNCKNSHCAQTVSVFKGIS 1155
Db 1056 ERKLCTPAIAIEALAGEVFETL-PKAYAEVLQCVVAAATTLSEHPETNAATGKLFKSSA 1114
Qy 1156 VNARQVRIEPPPKAPLGTVOQKRRQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQK 1215
Db 1115 IDAEVVVDMLCAMYDGRARDDATGQKPGKRSIAIAPSERVLGSTWKGCVTLLEHLQ 1174

Qy 1216 HKKKLRSPQIIVPTLFNLLSRCLEPLPQEOGNMEYTKQILILSCLLNICQKL---SPDGKG 1272
Db 1175 NKRKLANAQLVPLKFECLKYCLD--FEQSAVYVQIVQIVLALLLHICQKVEHSPAGP 1232
Qy 1273 IPKDIIDEEKFNVELIVQCIRLSEMPQTHHALLLGTVAGIFPDQKVLHNIMSTFTFWGA 1332
Db 1233 AP-PANASGIFKVGQIVQCIRGTQNPQTHHALLLAHVARVYVQVLLHNMEIFTFVGS 1291
Qy 1333 NVMLRDDTYSQVINKTVKMWIPALIOSDSGDSIEVSRNVEEIVVKLIISVFDALPHVPE 1392
Db 1292 SIM-----EQVVPILKIFSDILDVPE 1313
Qy 1393 HRRPILVOLVDTLGAKEFLMILLILPEQVVTVTVLAAAYGEK-----DAILEADTFW 1447
Db 1314 HRRILLYVKLLQTLGPAEYLMVFLGVLTIESVMKG-----GQKARNRPAKSRADAELA 1367
Qy 1448 FS-----VCCEPSVQHQISLMNIIQYLKLPKEEETIPKAVSNKSSQSEM 1496
Db 1368 GSELKMEVALLTAREPEPTVIIQCTCYLDYLDLPLMEIEK-----RKSDLGPME 1419
Qy 1497 LQV-----FNVETHSKQLRHFPLSVSMQSLSSNNPLKKVSVSGGPEIL--KGLBER 1549
Db 1420 LDVVDGSIFFNVKTHTRQLKHFYSVQVFGGLLSTVFVINKIAQNLNEETLGMKKYYQN 1479
Qy 1550 LLETVLGYISAVAQSMER---NADKLTVKWRALLSKAYDLDLKVALLPTETIPVIR 1605
Db 1480 LIVGILTYYNAVSKMLDKLEDGSDKIPLYWRAMLNCCYDILESTILLSADTLIIVH 1539
Qy 1606 GLVGNPLSPVRRKALDLNNKLOQNISW-KKTIVTRFLKLPDLLAIIVQRKKKEGEE--- 1661
Db 1540 GLLKHFRLMVRKVIELNNKLOYKQDYFNDSHYPGLLKLDPPLVELVQGLVEEQHVGT 1599
Qy 1662 ---EQAINQATYALYTLKLCNFGAENPDPFVPLVLTAV-KLIAPRKEEKVNLGSAALLCI 1718
Db 1600 AFERVMVIVQLSYIAVRHLKSKILTQYDTTKVQSVLAALMEELHGYKRNPNFTQLSSLICI 1659
Qy 1719 ABVTSTLALAIQPLPSLMPSLTTMKNTSELVSSEVY--LLSALAALQ-KVVETLPHFI 1775
Db 1660 GELCSHLGYSINFLPRFMPVWSKFLH--AQOQSGEPFDILTSSIVLLTKIVDILARFI 1717
Qy 1776 SPYLEGILSOVTHLEKITSEMGSAQANI--RLTSLKKTATTLAPRVLPAIKKTYKO- 1832
Db 1718 SPYLSMLVGLARLYAMIEQKDPRLGNMLSRLVLIWDSLTITITPRVLLPAIECYHTL 1777
Qy 1833 IEPKNWKNHMPMSILOEHIGXMKKEELTSHOSQLTAPFLEALDPRAQHSND---LEEV 1889
Db 1778 IGEGLSALGPLMGLLSTWFGMKSSSFDAIRSEVTELELTALQFRCYNSATDITYLDAI 1837
Qy 1890 GKTENCIIDCLVAMVVKLSEVTRFPLFKLPWAKTEDAPKDRLLTFYNLADCIKELKG 1949
Db 1838 DAAEAHVIFKAFVVLILKJSESTFRPLFYQVPEWSIRESSNDRAITIFNLCCHVAAALKS 1897
Qy 1950 LFTPLAGHLVKPFPADTLXQVNTSKTDEAPFDSND-----PEKCCLLQOILNCLY 2000
Db 1898 LFLVLFASDLVATVALLNATNSAKVGEHGGDGDGDAVAGELHFEVESKNVTLRLVYKLTLY 1957
Qy 2001 KIFLPTQHFISKERAXALMPLVDOLENRLGEEKFQBRVTKHLIPCLAQFSVMAADS 2060
Db 1958 SIVLYDNQNFINAVRFDMLLGPVSDHLEN---GLIVRVPVETRLVLDCLAQMAVAVMDS 2014
Qy 2061 LMKPLNYQILLKTRDSSPKVRFAALITVLAALAEKLENIVLLPESIPFLAEELMEDECEE 2120
Db 2015 LMRQLNYQVLMKTRNDAQVRLFALEACTEIAKRGESVAPLLPETIPFLAEELMEDDNE 2074
Qy 2121 VEHCQCKTIQOLETVLGEPLQSYF 2144
Db 2075 VEXAVHSCREIGRATGEDLOKYF 2098
RESULT 10
BP28_DROME
ID_BP28_DROME STANDARD; PRT; 2096 AA.

799	QY	ALKAPKSPKGIOWNPEQ-----LKEDSRDYHLHLIGLFEMWLNAGADAVHFRVLMKLFPI	853
800	Db	SEKADQ--PERQBWTRALQOSQLIILPEAQDRLEVLN-----	835
854	QY	KVHLEDVFOLEKFCFSVLWTYSGSLNPLNCSVKVTLQTOALYVGCAMLSSQKTOCKHOLA	913
836	Db	VFYFERLPBLWPRDSYA-----VFRLOQFIILEAVLNSPKSQIDCGLV	879
914	QY	SISPPVVTSLNLNGLSPKVEKRRAAACIOALSG--VASPFFYLIIDHLSKABEITSDA	971
880	Db	H-----VLRVANACGSPLOTLRVOAINILQLISNRKLVSHVEQLVRSLLQRKSELSMDHE	934
972	QY	YVIQDLATLFEELQREKKLQKSHKSETLKNLLSCVYSCPS--YIADKMLKVLOQVNGE	1028
935	Db	---QYALIIYTIILEPEKATAKERVLVSKURSVLALASDPKOSPICITASLAAALKHVND	991
1029	QY	MVLSQLLPMAEQILL-----EKIQKEP-----TAVUKDEAMV-LHLTLGKYNBF	1070
992	Db	NFLNELPLGLDLSKTIAGEDNQNIKQLPWPHESEYKSVIERFEGRVALNVLRRK--DL	1049
1071	QY	SVSLNEDPKSLDIFTKAVHTTKELVAGMPTIOITALEKITRPFPAAISDEKVQOK---	1126
1050	Db	AWKLFEDSPAQYDTYV-----QLEBQLQPLPCVLLNSLTPETF-----EQMHAKHTA	1097
1127	QY	LIRMLFDLLVNCNKNSHCAQTSSVFKGISVNAEQVRIELEPPDKAKPLGTVOQKRQKMQ	1186
1098	Db	LIKLIVESATNSN-----DSIFLASHLLKRCRLDCQP--LVPILEMANTKVKEKQ	1148
1187	QY	QKKSODLESVQ-EVGGSYQWQVTLIIELOHKKKLRSQILVPTLTFNLISRCLEPLPQEO	1245
1149	Db	PVKRRSVQATQLDLTSPYKQGMQTLLELLEHKQLVGAELIIPPLFELLQACIL--TWB	1206
1246	QY	GNMEYTKQILSLCLMNICOKLSPDGKIKPDIIDEEKFNVELIVQICIRLSEMPOTHHAL	1305
1207	Db	SAAEYPKQLTILSSLHCCQTAQAGVQLVK-AMPESFRIELVVQSERTRNPOTOCHAL	1265
1306	QY	LLLGTVAGIPDPKVLNHNIMSIFTMGANVMRLDDTYSFQVINKTVKVIIPALIQSDSGDS	1365
1266	Db	LFUTHCAGMPQOVLHKHIVEIFTFVGSTVARHDDAFSLHIHNWVESIIPILL-----	1318
1366	QY	IEVSRNVEIWKIISVVDALPHVEHRRPLPIVLQVDTLGAEKELWILLILLFQYVT	1425
1319	Db	--LNTGHNELVIVLKVFDICTDVPVRRRLPIYATLFRVLPEKPHLMQPLCIFE----	1372
1426	QY	KTVLAAAYGKDAILEADTFEWFSVCCFSSVQHO-----IQSLMNLIQYLLKLPEEKEE	1480
1373	Db	SQVILEQVPQKVSTDKSRLEF---ARELTLMFEDPTVALQTCIRLLDYLAKLPATKSSL	1428
1481	QY	IPKAVSPNKSESQEMLQOVNVETHTSKOLRHKPFJLSVSPMSQILLSNNFLKKVRESGGP	1540
1429	Db	SGSGSGSVLSTEQ---QLPDRVTRTPKQLRHVKYLLTMDPLSGTSSCNEWEKMKRDPDN	1484
1541	QY	EILKGLERLETLVGLYISAVAOSMERNAKLTV-KFWRALLSKAYDLIDKVNALLPETET	1599
1485	Db	ELLPYQEFILTK-LATVGVLNGALEAESTPSEKFWRVLANHAHDVLNATGLLAPQH	1543
1600	QY	FIPVIRGLVGNPLPFSVRKALDLNNKNLQONISW-KKTIYTRFLKLPDILAIVQ--RKK	1656
1544	Db	FISVITELLKXHDHVVYRAIKWMDLLVTVKLSPSSDYFQOSNAEHGVLPAQLQEIINGILEG	1603
1657	QY	KEGEEQAINROTALYTLKLCKNFGAENDDPPVPVLXTAVKLAPERKEKNVLGSALL	1716
1604	Db	SSNSAQAKLQOQTALHALQALLARHGRDYEIECKSSLATUTTKITKRANVPKAVGVNVL	1663
1717	QY	CIABVTSTLEALAIQPLPSLMPSSLITTMKNTSELVSSEVYLLSAL-----AAL	1764
1664	Db	TLVEICASLRAHALAQLPKFAPQL-----TELLKEQVHQMASLKQGPDYVCSTLVTAL	1716
1765	QY	QKVVEITPHFISPYLEGILSQVHLEKITSMSGASQANILRTLSLKT-----	1812
1717	Db	HKLFKALPLFLGPIVLDIIGLARI-----SVQLENQLQDKRQVILKOKLADWS	1768
1813	QY	-LATTLAPRVLLPAIKKTYKO-IEKNWKNHMGPFM-SILOEHIGXMKKEELTSHQSOLTA	1869

Db	1769	AVAQGV	EVRLVPS	CAKAFSS	LLSQAYDELGHLMQ	OLLQSVRHNSAAQLQVQDPLSE	1827															
Qy	1870	FFLEALD	FRAQHS	ENDEE	--VGKTE	NCIDCLVAMVVKLS	EVTPRPLPKFLFDWAKTE	1927														
Db	1829	LFLQALN	FRLOVR	GLQRLQ	LVSDVEAS	ITETFTV	WILKLS	SEFRPMYSRVHKWA--LES	1887													
Qy	1928	APKDRLL	TPYNLAD	CTAEKLG	FTFLFAGH	VKPPFAD	TLXQVNI	SKTDEAF	PDSENDPEK	1987												
Db	1888	TSRETRLT	TPVLLN	RTAEAK	LSLVLF	ASDFVED	SRLLTEH	RSIRPE	FVEVEREDVD--	1946												
Qy	1988	CCULLQ	FILNCL	YKIFL	FDTHQ	HSISKER	AAXALM	PLVDQLEN	RLG--GEEXFQ	ERTVKHLI	2046											
Db	1947	---	LLMAIL	NTLHH	FVLYC	SEDFIN	DRFNV	LMPLVN	QLEN	DLVLGNESL	QQVLSN--	2000										
Qy	2047	PCIAQFS	VAMAD	DSLWK	PLMYQ	ILLKTR	SSPKVR	FAALIT	VLALAK	LKENI	YIVLLPES	2106										
Db	2001	-CIAQ	FAVA--	TNDVM	WKQLNS	QVLLKTR	TNSNPE	VRILAF	NSCVAI	ARKLG	ESYAALLPET	2058										
Qy	2107	IPFLAEL	MEDECE	BEVEH	OCCOKT	TIQOLET	VLGE	PLQSY	2143													
Db	2059	VPTIAEL	LEDEH	QORV	EKNTR	TGTG	VOELET	ILGES	VQKY	2095												
RESULT 11																						
Q6AWSO	PRELIMINARY; PRT; 2096 AA.																					
ID	Q6AWSO																					
AC	Q6AWSO;																					
DT	25-OCT-2004	(TtEMBLrel. 28, Created)																				
DT	25-OCT-2004	(TtEMBLrel. 28, Last sequence update)																				
DT	25-OCT-2004	(TtEMBLrel. 28, Last annotation update)																				
DE	SD1791p.																					
GN	Name=CG10805;																					
OS	Drosophila melanogaster (Fruit fly).																					
OC	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;																					
OC	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;																					
OC	Ephydroidea; Drosophilidae; Drosophila.																					
OX	NCBI_TaxID=7227;																					
RN	[1]																					
RP	SEQUENCE FROM N.A.																					
RA	Stapleton M., Carlson J., Chavez C., Frise E., George R., Pacleab J.,																					
RA	Park S., Wan K., Yu C., Rubin G.M., Celniker S.;																					
RL	Submitted (AUG-2004) to the EMBL/GenBank/DBJ databases.																					
DR	EMBL; BT015178; AAT94407.1; -																					
DR	InterPro; IPR001395; Aldo/ket_red.																					
DR	InterPro; IPR008938; ARM.																					
DR	PROSITE; PS00063; ALDOXETO_REDUCTASE_3; UNKNOWN1.																					
QO	SEQUENCE 2096 AA; 237230 MW; 9CB27117C5DC4620 CRC64;																					

Query Match	19.0%;	Score 2051.5;	DB 2;	Length 2096;
Best Local Similarity	28.1%;	Pred. No. 6.9e-87;		
Matches 633: Conservative	423;	Mismatches 925;	Indels 273;	Gaps 65;

QY	2	TSLAAQQLORLALPOSDASLILSRDSVASLFLDPKAAATIDRDTAFPAICTGTGLEELLGLDPS	61
Db	3	TALAAQLQKLAAPQSSVTLADARSRASITLDPKKAATKDRSIIYEIGLTGTQOEITDFNPA	62
QY	62	FEQFEAPFLSQLAATLRSVOTKAVNKOLDENISLFLIHLSPPYLLKPAQKCLEWLHHRF	121
Db	63	FKFQTLTFDEATITLRSVELPEINKWLDAAIAKFLRLSPYLLLRPAHMAFEWLLRRF	122
QY	122	HHIYNODSLIACVLPYHETRI FVRVTQLLKINNSKRWFWMLPVKQSGVPLAKGTLITH	181
Db	123	QVHEYNRSEVNALLPYPHETMI FVOIVKTMELRSDDGWYRLRQLRPGVPLAKTAIINR	182
QY	182	CYKOLGFMDFICSLVTSKVFASYPGSSAQLRVLAFYASTIVSALVAADSDVSNIIAK	241
Db	183	AASNPAFLGFCISQTKAVK---ELGPRAHQLOQAQINFYATVVVYGALQTAKPLOQDWHIT	239
QY	242	LFPIYQKGLKSGSLPDYRAATYMIICQISVKVTMENTFVNSLASQIIKTLTKIPSLIKOGL	301
Db	240	ILESLLRGLIGDITDFMAAAVIVIAQLVSRTKLXSKVCNALLERVAN--CFPERLHSSGL	297

QY 302 SCLIVLQRPESLGGKPPHLCNVPDLITILHGISSETVDVSPLLRMLPHLVVSI-- 359
Db 298 LLLVCIIYQKQ-----QAALPHF--KPE--TILNLVGVKMWLISTSLAKGNIAIQSICM 347
QY 360 ----HHVTGETEGMDGQIYKRHLAEILTKISLKNLNDHLLASLLPREYIS----- 406
Db 348 PLMTGAVAAIRDDDDASSNSCKFLDNLLSEVPMKPTAQQLNCLFDTYVETAIADPEPM 407
QY 407 -----YSSQEMDSNKVSLNBOFLPIRILLESKYPTLTVLEELHKEIADL 454
Db 408 ETNSNEDDDTIVIDSDEIETKTT--FQAWYSTYLEKERRYPEAFDLSVKEALRSKST 466
QY 455 KQQLFHFQV---SLSTSGGKYQFLADSDTSLMLSLNHPLAPRILAMNH-LKKIMTKS 510
Db 467 SNKQKALKAALGFRLNTTDEKAHAYE-----KLYHSADWRLSAVQKLLQNLNLT 519
QY 511 EGVDESPIKEAVLARIGDDNDIVVLSAIS-AFEIFKEHFSSEVTISNLLNLFORAELSKN 569
Db 520 RESRVKLLQCLPRINDSGAVVSTLLSLPTEELAEMGLPPLAQTLCHLLYRAQSEKD 579
QY 570 GEWYEVUKIAADILIKBEIISENDQSLNQVVCILPFPVWINDDTESAEMKIAIYLSKSG 629
Db 580 EEWQPVVPLAVRHLTSALVSGSYD--TNLVLLALMPLLPFGEALAEHOKALRILLG-SD 636
QY 630 ICSLHPLLRGWEALENVIKSTPKGLIGVANQKMIELLA-DNINLGDPSS-SMLKWVEDL 687
Db 637 FVSKVPFLA--ELKVSNNKFDNF-----VGEHRQHFLDIIASSNQELSSQERALLQSVED- 689
QY 688 ISVGEESFNLKOKVTHVILSVLSCSSLSKETHPPFAIRVFSLLQKTKLESVITAV 747
Db 690 --HGGEIYQKASQLT-HLLLLITAYAKRELQPSHLMLEKGLYSRRLQ--FRVNGS 744
QY 748 E-----IPSEWHIELMDRGIPVELMAHYVEELNSTQRAVEDSVFLVSLKFKPI---Y 798
Db 745 QNTPCNAPLQLYVDFLLTL-VKNTKWT---ALASTPNQMTDELRLCLRLLEIICAQVF 799
QY 799 ALKAPKSPKCDIWNPEQ-----LKEDSRDYHLHLLIGLFEMMLNGADAVHFRVLMKLF 853
Db 800 SEKADQ--PERQWTRALQOSLOLILPEAQDRDLVLSN----- 835
QY 854 KVHLEDVOLFCKSVLWTVYSSLSNPLNCSVKTVLQTOALYVGCAMLSQKTOCKHQLA 913
Db 836 -----FYVFERLPELWPRDSYA-----VFLQGFILAEVLSNPKSQIDCGLV 879
QY 914 STSPVVTSLNLNGLSPVKEVRRAAIOCLQALSG--VASPFYLIIDHLISKABEITSDA 971
Db 880 H-----VLRVANAGSPQLTLRQVAINTLQIISNRKLVSHEVQLVRSLLQKSELSNDHE 934
QY 972 YVIODLATLPEELQREKKLKHOKLSETLKNLLSCVVCPS-----YIAKOLMKVLQVNGE 1028
Db 935 ---QYALILYILEPEKATAKERVLVLSKLKRSVLALASDPKQSPICITASLLAALKHVND 991
QY 1029 MVLSQLPMAEQLEKIQEPTAVLXDEAMVHLTLGKYNFSPVSLN--EDPKSIDIFI 1086
Db 992 NFLNELPLGLDSLKTI-----TAGEDNQIKHLP-WPHSEIYKSVIERPEGRVAINVLL 1045
QY 1087 K-----AVHTTKELYAGMPTIOITALEKITKPFPAIAISDEKVOOK-----LLR 1129
Db 1046 RKDLAWKLPEDSPAQDYTVQLSQKLPCLVLLNSITPETF-----EQMHAGKIALIK 1100
QY 1130 MLFDLLVKNKNSHCAQTVSVFKGISVNAEQVRIELEPPDKAKPLGTQVQKRRQKQOKK 1189
Db 1101 LIVESATNSDN-----DSIFLASHRLKRCRLDQCP--LVPILLEMANTKVEKKQPVK 1151
QY 1190 SQDLSEVQ--EVGGSYQWQVTLILLELQHKKLRSPOILVPTLFLNLSRCLLEPLPOQSGNM 1248
Db 1152 RRSVQATQDLTSPYKWKQMTLLELLEHKKQLLGAELLIPPLPELLQACL--TMEHSAA 1209
QY 1249 EYTKQILSLCLNICOKLSPDGKIPKDIIDDEKFNVELIVQICIRISEMPTQTHHALLLL 1308
Db 1210 EYPKQILLSLLHCCCAQASAGVQVK--AMPESFRIELVQSLRNTRNPQTQOHALFL 1268
QY 1309 GTVAGIFDPKVLNINMISITFTFMGANVMRLDDTVSFQVINKTVKQVIPALIQSDSGDSIEV 1368

Db 1269 THCAGMPYQOVLKHKVIEFTFVGSTVARHDDAFSLHIIHNWVESIIPILL-----L 1319
QY 1369 SRNVEELVKVILSVFVDALPHVPEHRELPIVLQVDTLGAELFWILLILLPEQYVTKV 1428
Db 1320 NTGHNELVILPVLKVFADICTDVPVRRRLPLYATLFRVLEPKHELMQFLCIFE-----SQV 1375
QY 1429 LAAYGEKDAILBEADTEFWFSVCESFVHQ-----IQSLMNILQYLKLPKEEKEETIPK 1483
Db 1376 LLEQVPQKVTSDKSRDLDF-----ARELTLMFEDPTVAJOTCIRLLDYLAKLPAKSSLSGG 1431
QY 1484 AVSFNKSESQEMLOVPNVETHTSKQLRHFKFLSVSPMSQSSLLSNFLLKVVESGGPIL 1543
Db 1432 SGSSVLSTEQ---QLFDVTRTRFKQLRHYKYLIMDFLSGISSCNEKWKMRDPDNELL 1487
QY 1544 KGLEERLELTVLGVISAVASOMERNADKLTV-KFWRALLSKAYDOLLQKVNALLPTETFIP 1602
Db 1488 PYQEFILKT-LAYGVINGALAEASSETPSLEKFRVRLANHAHDVLDONAIGLLAPQHFLS 1546
QY 1603 VIRGLVGNPLPSVRRKALDILLNNKLOONIISW-KKTIIVTRFKLVPDILAIQV--RKKKEG 1659
Db 1547 VITELLKHDHVYVRIKVMDDLVTKLSPSTDYFQOSNAEHFGVLPAEQEIINGILQSSN 1606
QY 1660 EEEQAINRQATLYTLKLLCKNFGAENPDPPVPLVXTAVKLIAPERKEEKVLSGALICIA 1719
Db 1607 SAQAKLQOQTALHALQLAFRHRGRDYIECRSLATLTITKRRANVPKAVGVNVVLTV 1666
QY 1720 EVTSTLEALAIPOPLSIMPSSLTMTKNTSELVSSEVLLSAL-----AALQKV 1767
Db 1667 EICASKAHALAOQKPAFOL-----TELLKEQVHOMASLKQGGPDYVGVSTLVTALHKL 1719
QY 1768 VETLPHFISPLEGILSQVHLEKITSEMGSASQANIRLTSKKT-----LA 1814
Db 1720 FKALPLFLGPLYVDIIGGLARL-----SVQLENPQLQDKRTQVLKQKLADVNSAVA 1771
QY 1815 TTLAPRVLLPAIKTKYQ--IEKNKWHMGPFM-SILOEHIGXMKKBELTSHQSOLTAFFL 1872
Db 1772 QGVEVRILVPSKAKAFSLLEQQAQYDELGHLMQOQLLQSVRHNSAAQLOQPVDPLSELFL 1831
QY 1873 EALDFFRAQHSENDLEB--VGKTENCIIDCLVAMVVKLSEVTFEPLFKFLFDWAKTDEAPK 1930
Db 1832 QALNFRLOVRGLQQLQVSDVEASTETFTVWLKLSSETSPRPMYSRVHKA--LESTSR 1890
QY 1931 DRLLTFYNLADCAIEAKLGLFTLFAHVLKVPFADTLXQVNIISKTDEAFFDSENDPEKCC 1990
Db 1891 ETRLTYPLLTNRIAEALKSLFVLFSDFVEDSSRLTTEHNSIRPEFEVEREDVD---- 1946
QY 1991 LLOFILNCLYKIFLFTQHFISKERAXALMPLVDQLENRLG--GEBKQBRVTKHLIPCI 2049
Db 1947 LLMAILNTLHHVFLYCSDFINDHRENVLMPPLVNQLENDLVLGNESLQVLSN-----CI 2002
QY 2050 AOPSVAMADDSLWKPLANYQILLKTRDSSPKVRFAALITVLALAEKLENVILLPSIPI 2109
Db 2003 AQFAVA-TNDVMWKQLNSQVLLKTRTSNPEVRIAFNSCVAIARKIGESYAALLPETVPF 2061
QY 2110 LAELMEDECEBEVHCQCKTIQOQLETVLGEPLQSY 2143
Db 2062 IAELEDEHORVEKTRTGVQLETLIGESVQKY 2095
RESULT 12
Q8CCT5 PRELIMINARY; PRT; 408 AA.
AC Q8CCT5;
DT 01-MAR-2003 (Tremblrel. 23, Created)
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
DE 01-MAR-2004 (Tremblrel. 26, Last annotation update)
DE Mus musculus adult male olfactory brain cDNA, RIKEN full-length
DE enriched library, clone:6430400D06 product:hypothetical ARM repeat
DE structure containing protein, full insert sequence.
GN Name:B13001612Rik; Synonyms=BC019693;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

DR Pfam; PF02985; HEAT; 1.
 SQ SEQUENCE 349 AA; 39921 MW; 3A359597FF7F7079BB CRC64;
 Query Match 16.5%; Score 1779; DB 2; Length 349;
 Best Local Similarity 99.1%; Pred. No. 3e-75;
 Matches 346; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1796 MGSASQANIRLTSLSKKTATLAPRVLLPAIKTKYQIERNKWNHMGPFMSILQEHIGXM 1855
 Db 1 MGSASQANIRLTSLSKKTATLAPRVLLPAIKTKYQIERNKWNHMGPFMSILQEHIGXM 60
 QY 1856 KKEELTSHSQQLTAFFLEALDFAQHSNDLEEVGKTENCII DCLVAMVVKLSVTFRPL 1915
 Db 61 KKEELTSHSQQLTAFFLEALDFAQHSNDLEEVGKTENCII DCLVAMVVKLSVTFRPL 120
 QY 1916 FFKLFDWAKTEDAPKORLLTFYNLADCI AEKLGFLTFAGHLVKPFPADTLQVNISKTD 1975
 Db 121 FFKLFDWAKTEDAPKORLLTFYNLADCI AEKLGFLTFAGHLVKPFPADTLQVNISKTD 180
 QY 1976 EAPFDSNDPEKCCLLQFIINCLYKIFLFDTOHFISKERAXALMPLVDQLENRLGEE 2035
 Db 181 EAPFDSNDPEKCCLLQFIINCLYKIFLFDTOHFISKERAXALMPLVDQLENRLGEE 240
 QY 2036 KFOERVTKHLIPCI AQFSVAMADSLWKPLNYQILLKTRDSSPKVRFPAALITVLAAEKL 2095
 Db 241 KFOERVTKHLIPCI AQFSVAMADSLWKPLNYQILLKTRDSSPKVRFPAALITVLAAEKL 300
 QY 2096 KENYIVLLPESIPFLAEIMEDECEVEHOCQKTIQOLETVLGEPLOSIF 2144
 Db 301 KENYIVLLPESIPFLAEIMEDECEVEHOCQKTIQOLETVLGEPLOSIF 349

RESULT 14
 Q8VCK1 PRELIMINARY; PRT; 349 AA.
 ID Q8VCK1
 AC Q8VCK1
 DT 01-MAR-2002 (TEMBLrel. 20, Created)
 DT 01-MAR-2002 (TEMBLrel. 20, Last sequence update)
 DT 01-MAR-2004 (TEMBLrel. 26, Last annotation update)
 DE B13001612Rik protein.
 GN Name=B13001612Rik;
 OS Mus musculus (Mouse)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Mix FVB/N; TISSUE=Mammary tumor;
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.L., Feingold B.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Sherman C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.P., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.D., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
 RA Jones S.J., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Mix FVB/N; TISSUE=Mammary tumor;
 RX Director MGC Project;

RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC019693; AH19693.1; -
 DR MGI; MGI:2442524; B13001612Rik.
 SQ SEQUENCE 349 AA; 40166 MW; 9763D0331AD0F515 CRC64;
 Query Match 15.6%; Score 1682; DB 2; Length 349;
 Best Local Similarity 92.8%; Pred. No. 1e-70;
 Matches 324; Conservative 14; Mismatches 11; Indels 0; Gaps 0;

QY 1796 MGSASQANIRLTSLSKKTATLAPRVLLPAIKTKYQIERNKWNHMGPFMSILQEHIGXM 1855
 Db 1 MGSASQANIRLTSLSKKTATLAPRVLLPAIKTKYQIERNKWNHMGPFMSILQEHIGXM 60
 QY 1856 KKEELTSHSQQLTAFFLEALDFAQHSNDLEEVGKTENCII DCLVAMVVKLSVTFRPL 1915
 Db 61 KKEELTSHSQQLTAFFLEALDFAQHSNDLEEVGKTENCII DCLVAMVVKLSVTFRPL 120
 QY 1916 FFKLFDWAKTEDAPKORLLTFYNLADCI AEKLGFLTFAGHLVKPFPADTLQVNISKTD 1975
 Db 121 FFKLFDWAKTEDAPKORLLTFYNLADCI AEKLGFLTFAGHLVKPFPADTLQVNISKTD 180
 QY 1976 EAPFDSNDPEKCCLLQFIINCLYKIFLFDTOHFISKERAXALMPLVDQLENRLGEE 2035
 Db 181 EAPFDSNDPEKCCLLQFIINCLYKIFLFDTOHFISKERAXALMPLVDQLENRLGEE 240
 QY 2036 KFOERVTKHLIPCI AQFSVAMADSLWKPLNYQILLKTRDSSPKVRFPAALITVLAAEKL 2095
 Db 241 KFOERVTKHLIPCI AQFSVAMADSLWKPLNYQILLKTRDSSPKVRFPAALITVLAAEKL 300
 QY 2096 KENYIVLLPESIPFLAEIMEDECEVEHOCQKTIQOLETVLGEPLOSIF 2144
 Db 301 KENYIVLLPESIPFLAEIMEDECEVEHOCQKTIQOLETVLGEPLOSIF 349

RESULT 15
 Q8T9E7 PRELIMINARY; PRT; 1690 AA.
 ID Q8T9E7
 AC Q8T9E7
 DT 01-JUN-2002 (TEMBLrel. 21, Created)
 DT 01-JUN-2002 (TEMBLrel. 21, Last sequence update)
 DT 01-MAR-2004 (TEMBLrel. 26, Last annotation update)
 DE SD03723P.
 GN ORFNames=CG10805;
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Stapleton M., Brokstein P., Hong L., Agbavani A., Carlson J.,
 RA Champe M., Chavez C., Dorsett V., Farfan D., Friese E., George R.,
 RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
 RA Nunco J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,
 RA Yu C., Lewis S.E., Rubin G.M., Celniker S.;
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY069800; RAL39945.1; -
 DR FlyBase; FBgn0031864; CG10805.
 DR InterPro; IPR008938; ARM.
 SQ SEQUENCE 1690 AA; 191874 MW; B7CE254A4FBADF09 CRC64;
 Query Match 13.5%; Score 1462.5; DB 2; Length 1690;
 Best Local Similarity 27.1%; Pred. No. 1.6e-59;
 Matches 496; Conservative 342; Mismatches 755; Indels 237; Gaps 59;

QY 402 EYISYVSQSEMDSNKVSLLNEQFLIRLLSKYPRTLQVLEHKLKADKKQELPH 461
 Db 9 DDTIVIDSDEIETKTT-FQAWYSTYLEKERRYPEAFDLSVKEALRSKSTSNRQAL 67
 QY 462 QFV---SLSTSGKYQFLADSDTSLMLSLNHPPLAPVIRILAMNH-LKKIMKTSKEGVDSF 517
 Db 68 KLALGFRUNTTDEKAKHAYE-----KLYHSADWRUSAVQKLQNLNVTKRERSVKL 120

